

(19)日本国特許庁 (J P)

(12) 公 開 特 許 公 報 (A)

(11)特許出願公開番号

特開平11-154146

(43)公開日 平成11年(1999)6月8日

(51)Int.Cl.⁶
G 0 6 F 15/18
G 1 0 L 3/00
// G 0 6 T 7/00

識別記号
5 6 0
5 2 1
5 3 1

F I
G 0 6 F 15/18
G 1 0 L 3/00
G 0 6 F 15/70
5 6 0 A
5 2 1 F
5 3 1 E
4 6 5 A

審査請求 未請求 請求項の数20 O L (全 22 頁)

(21)出願番号 特願平9-322883
(22)出願日 平成9年(1997)11月25日

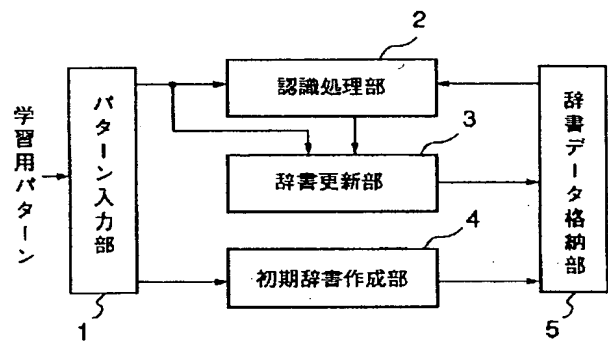
(71)出願人 000003078
株式会社東芝
神奈川県川崎市幸区堀川町72番地
(72)発明者 黒沢 由明
神奈川県川崎市幸区小向東芝町1番地 株
式会社東芝研究開発センター内
(74)代理人 弁理士 鈴江 武彦 (外6名)

(54)【発明の名称】 パターン認識方法および記録媒体

(57)【要約】

【課題】理論的に方式全体の整合が取れ、高速学習可能なパターン認識辞書作成方法およびそれにより作成された辞書を用いた精度の高いパターン認識が行えるパターン認識方法を提供する。

【解決手段】入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことによりパターン認識辞書を作成するパターン認識辞書作成方法において、前記更新処理は、前記パターンベクトルに行列を掛け、さらに該パターンベクトルの転置を掛ける計算式を含む各カテゴリに対する該ベクトルの類似度を算出するための類似度計算式に従った認識結果に基づく評価式の値を最適化する方向に前記行列を更新する漸化式に従って、カテゴリ毎の前記行列を更新することを特徴とする。



【特許請求の範囲】

【請求項1】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことにより作成されたパターン認識辞書を用いて、入力されたパターンを認識するパターン認識方法において、

前記更新処理は、前記パターンベクトルに行列を掛け、さらに該パターンベクトルの転置を掛ける計算式を含む各カテゴリに対する該ベクトルの類似度を算出するための類似度計算式に従った認識結果に基づく評価式の値を最適化する方向に前記行列を更新する漸化式に従って、カテゴリ毎の前記行列を更新することを特徴とするパターン認識方法。

【請求項2】 前記カテゴリ毎の行列の初期行列は、前記パターンベクトルの集合に基づき算出された第1の行列と、各カテゴリに分類された前記パターンベクトルの集合に基づき算出された各カテゴリ毎の第2の行列とから算出される行列であることを特徴とする請求項1記載のパターン認識方法。

【請求項3】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことにより作成されたパターン認識辞書を用いて、入力されたパターンを認識するパターン認識方法において、

前記更新処理は、前記パターンベクトルに第1の行列を掛け、さらに該パターンベクトルの転置を掛ける第1の計算式と前記第1の行列の固有ベクトルからなる第2の行列とに基づく各カテゴリに対する該パターンベクトルの類似度を計算するための類似度計算式に従った認識結果に基づく評価式の値を最適化する方向に前記第1の行列を更新する漸化式に従って、カテゴリ毎の前記固有ベクトルを含む前記第1の行列を更新し、前記漸化式には、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための関数で、前記正解カテゴリと不正解カテゴリの類似度の差が等しい部分で高い値を取り、前記正解カテゴリと不正解カテゴリの類似度の差が異なる部分で低い値をとる窓関数を有することを特徴とするパターン認識方法。

【請求項4】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことにより作成されたパターン認識辞書を用いて、入力されたパターンを認識するパターン認識方法において、

前記更新処理は、前記パターンベクトルに第1のベクトルを減算して生成された第2のベクトルに第1の行列を掛け、さらに第2のベクトルの転置を掛ける第1の計算式と、前記第1の行列の固有ベクトルから前記第1のベクトルを減算したベクトルとに基づく前記パターンベクトルの相違度を計算するための相違度計算式に従った認

識結果に基づく評価式の値を最適化する方向に前記第1の行列を更新するための第1の漸化式および前記第1のベクトルを更新するための第2の漸化式に従って、カテゴリ毎の前記固有ベクトルおよび前記カテゴリ毎の第1のベクトルのうちの少なくとも一方を更新することを特徴とするパターン認識方法。

【請求項5】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことにより作成されたパターン認識辞書を用いて、入力されたパターンを認識するパターン認識方法において、

前記更新処理は、前記パターンベクトルと第1のベクトルとの内積の2乗項に第1の係数を乗算し、さらに第2の係数を加算する第1の計算式を少なくとも含む各カテゴリに対する前記パターンベクトルの類似度を算出するための類似度計算式に従った認識結果に基づく評価式の値を最適化する方向に前記第1のベクトルを更新するための漸化式に従ってカテゴリ毎の前記第1のベクトルを更新することを特徴とするパターン認識方法。

【請求項6】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことにより作成されたパターン認識辞書を用いて、入力されたパターンを認識するパターン認識方法において、

前記更新処理は、前記パターンベクトルと予め定められた制限のある辞書データとの類似度計算式に従った認識結果に基づく第1の評価式に、さらに前記辞書データの予め定められた制限を表した式に任意の定数を乗じたものを加算した第2の評価式を最適化する方向に前記辞書データを更新するための漸化式に従ってカテゴリ毎の前記辞書データを更新することを特徴とするパターン認識方法。

【請求項7】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことにより作成されたパターン認識辞書を用いて、入力されたパターンを認識するパターン認識方法において、

前記更新処理は、前記パターンベクトルと辞書データとの類似度計算式に従った認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従ってカテゴリ毎の前記辞書データを更新し、前記漸化式には、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための関数で、前記正解カテゴリと不正解カテゴリの類似度の差が等しい部分で高い値を取り、前記正解カテゴリと不正解カテゴリの類似度の差が異なる部分で低い値をとる窓関数を有し、前記辞書データを更新する際、前記正解カテゴリと不正解カテゴリの類似度の差が所定の範囲内の時のみに前記窓関数を用いることを特徴とするパターン認識辞書作成方法。

【請求項 8】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことにより作成されたパターン認識辞書を用いて、入力されたパターンを認識するパターン認識方法において、

前記更新処理は、前記パターンベクトルと辞書データとの類似度計算式に従った認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従ってカテゴリ毎の前記辞書データを更新し、前記漸化式には、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための窓関数を有し、この窓関数の特性を定めるパラメータ値は、前記パターンベクトルの複数の認識結果を比較することにより設定されることを特徴とするパターン認識方法。

【請求項 9】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことにより作成されたパターン認識辞書を用いて、入力されたパターンを認識するパターン認識方法において、

前記更新処理は、前記パターンベクトルと辞書データとの類似度計算式に従った認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従ってカテゴリ毎の前記辞書データを更新し、前記漸化式には、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための窓関数を有し、この窓関数の特性を定めるパラメータ値は、前記パターンベクトルの認識結果に基づく前記正解カテゴリと不正解カテゴリの類似度に関する頻度分布に基づき決定されることを特徴とするパターン認識方法。

【請求項 10】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新処理を繰り返すことにより作成された大分類パターン認識辞書を用いて、入力されたパターンを認識するパターン認識方法において、

前記更新処理は、前記パターンベクトルと辞書データとの類似度計算式に従った該パターンベクトルの大分類の認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従って正解カテゴリと不正解カテゴリの前記辞書データを更新することを特徴とするパターン認識方法。

【請求項 11】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新を繰り返すことによりパターン認識辞書を作成するためのプログラムを記録した機械読み取り可能な記録媒体であって、

前記パターンベクトルに行列を掛け、さらに該パターンベクトルの転置を掛ける計算式を含む各カテゴリに対する該ベクトルの類似度を算出するための類似度計算式に

従った認識結果に基づく評価式の値を最適化する方向に前記行列を更新する漸化式に従って、カテゴリ毎の前記行列を更新させる更新手段を実行するプログラムを記録した記録媒体。

【請求項 12】 前記パターンベクトルの集合に基づき算出された第 1 の行列と、各カテゴリに分類された前記パターンベクトルの集合に基づき算出された各カテゴリ毎の第 2 の行列とから前記カテゴリ毎の行列の初期行列を算出させる算出手段を実行するプログラムをさらに記憶した請求項 11 記載の記録媒体。

【請求項 13】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新を繰り返すことによりパターン認識辞書を作成するためのプログラムを記録した機械読み取り可能な記録媒体であって、

前記パターンベクトルに第 1 の行列を掛け、さらに該パターンベクトルの転置を掛ける第 1 の計算式と前記第 1 の行列の固有ベクトルからなる第 2 の行列とに基づく各カテゴリに対する該パターンベクトルの類似度を計算するための類似度計算式に従った認識結果に基づく評価式の値を最適化する方向に前記第 1 の行列を更新する漸化式に従って、カテゴリ毎の前記固有ベクトルを含む前記第 1 の行列を更新させる更新手段を実行するプログラムを記録し、前記漸化式には、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための関数で、前記正解カテゴリと不正解カテゴリの類似度の差が等しい部分で高い値を取り、前記正解カテゴリと不正解カテゴリの類似度の差が異なる部分で低い値をとる窓関数を有することを特徴とする記憶媒体。

【請求項 14】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新を繰り返すことによりパターン認識辞書を作成するためのプログラムを記録した機械読み取り可能な記録媒体であって、

前記パターンベクトルに第 1 のベクトルを減算して生成された第 2 のベクトルに第 1 の行列を掛け、さらに第 2 のベクトルの転置を掛ける第 1 の計算式と、前記第 1 の行列の固有ベクトルから前記第 1 のベクトルを減算したベクトルとに基づく前記パターンベクトルの相違度を計算するための相違度計算式に従った認識結果に基づく評価式の値を最適化する方向に前記第 1 の行列を更新するための第 1 の漸化式および前記第 1 のベクトルを更新するための第 2 の漸化式に従って、カテゴリ毎の前記固有ベクトルおよび前記カテゴリ毎の第 1 のベクトルのうちの少なくとも一方を更新させる更新手段を実行するプログラムを記録した記録媒体。

【請求項 15】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新を繰り返すことによりパターン認識辞書を作成するための

プログラムを記録した機械読み取り可能な記録媒体であって、

前記パターンベクトルと第1のベクトルとの内積の2乗項に第1の係数を乗算し、さらに第2の係数を加算する第1の計算式を少なくとも含む各カテゴリに対する前記パターンベクトルの類似度を算出するための類似度計算式に従った認識結果に基づく評価式の値を最適化する方向に前記第1のベクトルを更新するための漸化式に従ってカテゴリ毎の前記第1のベクトルを更新させる更新手段を実行するプログラムを記録した記録媒体。

【請求項16】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新を繰り返すことによりパターン認識辞書を作成するためのプログラムを記録した機械読み取り可能な記録媒体であって、

前記パターンベクトルと予め定められた制限のある辞書データとの類似度計算式に従った認識結果に基づく第1の評価式に、さらに前記辞書データの予め定められた制限を表した式に任意の定数を乗じたものを加算した第2の評価式を最適化する方向に前記辞書データを更新するための漸化式に従ってカテゴリ毎の前記辞書データを更新させる更新手段を実行するプログラムを記録した記録媒体。

【請求項17】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新を繰り返すことによりパターン認識辞書を作成するためのプログラムを記録した機械読み取り可能な記録媒体であって、

前記パターンベクトルと辞書データとの類似度計算式に従った認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従ってカテゴリ毎の前記辞書データを更新させる更新手段と、

前記辞書データを更新する際、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための関数で、前記正解カテゴリと不正解カテゴリの類似度の差が等しい部分で高い値を取り、前記正解カテゴリと不正解カテゴリの類似度の差が異なる部分で低い値をとる窓関数を、前記正解カテゴリと不正解カテゴリの類似度の差が所定の範囲内の時のみに前記漸化式に適用させる手段と、
40 50
を実行するプログラムを記録した記録媒体。

【請求項18】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新を繰り返すことによりパターン認識辞書を作成するためのプログラムを記録した機械読み取り可能な記録媒体であって、

前記パターンベクトルと辞書データとの類似度計算式に従った認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従ってカテゴリ毎の前記辞書データを更新させる更新手段と、

前記漸化式の有する、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための窓関数の特性を定めるパラメータ値を、前記パターンベクトルの複数の認識結果を比較することにより設定させる設定手段と、
50
を実行するプログラムを記録した記録媒体。

【請求項19】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新を繰り返すことによりパターン認識辞書を作成するためのプログラムを記録した機械読み取り可能な記録媒体であって、

前記パターンベクトルと辞書データとの類似度計算式に従った認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従ってカテゴリ毎の前記辞書データを更新させる更新手段と、

前記漸化式の有する、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための窓関数の特性を定めるパラメータ値を、前記パターンベクトルの認識結果に基づく前記正解カテゴリと不正解カテゴリの類似度に関する頻度分布に基づき設定させる設定手段と、
20
を実行するプログラムを記録した記憶媒体。

【請求項20】 入力されたパターンベクトルに基づきパターン認識辞書に登録されている辞書データの更新を繰り返すことにより大分類パターン認識辞書を作成するためのプログラムを記録した機械読み取り可能な記録媒体であって、

前記パターンベクトルと辞書データとの類似度計算式に従った該パターンベクトルの大分類の認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従って正解カテゴリと不正解カテゴリの前記辞書データを更新させる更新手段を実行するプログラムを記録した記憶媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明はパターン認識分野における辞書作成方法に関する。

【0002】

【従来の技術】 文献「The Self-Organizing Map」(Proceedings of the IEEE (1990))には、LVQ(学習ベクトル量子化)が紹介されている。これは入力ベクトルを x とし、それぞれのカテゴリに対して m を参照ベクトルとし、相違度 s を $s = \|x - m\|$ と定義するもので、各カテゴリ中で最小の相違度を持つカテゴリを入力ベクトル x の認識結果として出力する。

【0003】 この時、 m の学習には次の手順を使う。学習パターン x を認識した時、認識結果とし得られたカテゴリを c 、参照ベクトル m_c として、

正解のとき、 $mc' = mc + \alpha (x - mc)$... (61)

不正解のとき、 $mc' = mc - \alpha (x - mc)$... (62)

のようにして辞書を更新する。上記以外の方法にLVQ2、LVQ3があり、認識結果として得られる相違度に基づいて、式(61)と式(62)を使いわけている。

【0004】また、GPDと呼ばれる考え方をユークリッド距離に適用する方式も知られている。R(x)をxを認識した時に得られる認識結果とし、 $\alpha(R(x))$ をR(x)によって決まる係数として、一般的に、 $mc' = mc + \alpha(R(x))(x - mc)$ を更新ルールとするものである。

【0005】これらの方式では相違度がユークリッド距離で単純であり、認識精度に問題があった。また、 α を最適に決めることも難しい課題となっており実用上、問題があった。

【0006】文献「パターン認識と部分空間法」(産業図書(1986))にはLSM(学習部分空間法)が紹介されている。これは、入力ベクトルをxとし、それぞれのカテゴリに対してPをプロジェクションとし、類似度sを $s = x^T P x$ と定義するものである。なお、 x^T は、xの転置ベクトルである。そして、各カテゴリ中で最大の類似度を持つカテゴリをxの認識結果として出力する。

【0007】Pを構成する正規直交ベクトルを ψ_i とした時、Gを ψ_i をr個並べた行列として $P = G G^T$ と表せるものとする。R(x)をxを認識した時に得られる認識結果として、

$$\psi_i' = \psi_i + \alpha(R(x))(x, \psi_i)x$$

とし、 ψ_i をたとえばシュミットの直交化法で正規直交化して新しい ψ_i を求めることにより学習を行う。

【0008】平均ベクトルmを持つ形式で、 $s = \|x - m\|^2 - (x - m)^T P (x - m)$

も投影距離として知られているが、 ψ_i については上と同様に求め、また、mについて、 $m' = m + 2\alpha(R(x))((x - m) - P(x - m))$ によって更新する方式も知られている。

【0009】また、同じ文献に紹介されているALSM(平均学習部分空間法)や特開昭第56-137483号に記載されている辞書作成方法は、次のようなものである。

【0010】Kを正値対称行列として、Kの固有ベクトルの一部を用いてPを構成し、類似度を $s = x^T P x$ と定義する。学習は認識が不正解であった場合に、正解カテゴリに対して $K' = K + \alpha x x^T$ 、不正解カテゴリに対して $K' = K - \beta x x^T$ によってKを更新することによって行う。

【0011】これらの方法では、理論的に何を最適化しようとしているものが不明確であったり、また不適當であって学習がうまくいかずに認識精度に問題を生じるこ

とがあった。

【0012】これに対して理論的に最適化しようとしているものが明確な方式は、部分空間法にGPDを適用したもので、文献「Discriminative Metric Design for Pattern Recognition」(ICASSP-95, Vol. 1, 5)に紹介されている。この方式は類似度 $s = x^T P x$ のPを別の複数の回転を表す行列Q_iの積として表し、このQ_iを学習するように構成されている。この方式ではPとQ_iの変換を毎回行う必要があり、学習時間が増大するという問題があった。

【0013】

【発明が解決しようとする課題】本発明は、上記問題点に鑑みなされたもので、パターン認識分野における辞書作成方法において、理論的に方式全体の整合が取れ、高速学習可能で、精度の高い辞書作成が行えるパターン認識辞書作成方法およびそれにより作成された辞書を用いた精度の高いパターン認識が行えるパターン認識方法を提供することを目的とする。

【0014】

【課題を解決するための手段】(1)本発明のパターン認識方法(請求項1、特性核学習)で用いられる前記更新処理は、入力パターンベクトルxに行列Kを掛け、さらに該入力パターンベクトルの転置 x^T を掛ける計算式を含む各カテゴリに対する該ベクトルの類似度を算出するための類似度計算式 $s = x^T K x$ に従った認識結果に基づく評価式(式(6))の値Jを最適化する方向に該行列Kを更新する漸化式(式(13))に従って、カテゴリ毎の前記行列Kを更新することより、高精度のパターン認識が可能なパターン認識辞書が高速に作成できる。

【0015】(2)本発明のパターン認識方法(請求項2、請求項1の特性核学習における初期行列の作成方法)で用いられる前記更新処理において、前記カテゴリ毎の行列の初期行列(K_0)は、前記パターンベクトルの集合に基づき算出された第1の行列(Q)と、各カテゴリに分類された前記パターンベクトルの集合に基づき算出された各カテゴリ毎の第2の行列(P)とから算出される行列である。

【0016】(3)本発明のパターン認識方法(請求項3、ALSMを用いた特性核学習：ウィンドウ関数を用いる場合)で用いられる前記更新処理は、前記パターンベクトル(x)に第1の行列(K)を掛け、さらに該パターンベクトルの転置を掛ける第1の計算式($x^T K x$)と前記第1の行列の固有ベクトル(ψ_i)からなる第2の行列(P)とに基づく各カテゴリに対する該パターンベクトルの類似度を計算するための類似度計算式($s = x^T P x$)に従った認識結果に基づく評価式(式

(6)) の値を最適化する方向に前記第1の行列を更新する漸化式(式(12))に従って、カテゴリ毎の前記固有ベクトルを含む前記第1の行列を更新し、前記漸化式には、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための関数で、前記正解カテゴリと不正解カテゴリの類似度の差が等しい部分で高い値を取り、前記正解カテゴリと不正解カテゴリの類似度の差が異なる部分で低い値をとる窓関数 $w(d)$ (式(8))を有することにより、高精度のパターン認識が可能なパターン認識辞書が高速に作成できる。

【0017】(4) 本発明のパターン認識方法(請求項4、ALS Mを用いた特性核学習に定数、平均ベクトルの学習を導入)で用いられる前記更新処理は、前記パターンベクトル (x) に第1のベクトル (m) を減算して生成された第2のベクトル $(x-m)$ に第1の行列

(K) を掛け、さらに第2のベクトルの転置を掛ける第1の計算式 $((x-m)^T K (x-m))$ と、前記第1の行列の固有ベクトル (ψ_i) から前記第1のベクトルを減算したベクトルとに基づく前記パターンベクトルの相違度を計算するための相違度計算式(式(40))に従った認識結果に基づく評価式(式(6))の値 J を最適化する方向に前記第1の行列を更新するための第1の漸化式(式(41))および前記第1のベクトルを更新するための第2の漸化式(式(42))に従って、カテゴリ毎の前記固有ベクトルおよび前記カテゴリ毎の第1のベクトルのうちの少なくとも一方を更新することにより、高精度のパターン認識が可能なパターン認識辞書が高速に作成できる。

【0018】(5) 本発明のパターン認識方法(請求項5、固有ベクトルの学習:2次形式評価式タイプで係数、定数学習を含むもの)で用いられる前記更新処理は、前記パターンベクトル (x) と第1のベクトル (ψ_i) との内積の2乗項に第1の係数 (a_i) を乗算し、さらに第2の係数 (c) を加算する第1の計算式を少なくとも含む前記パターンベクトルの類似度を計算するための類似度計算式(式(45)~式(46))に従った認識結果に基づく評価式(式(48))の値を最適化する方向に前記第1のベクトルを更新するための漸化式(式(47))に従ってカテゴリ毎の前記第2のベクトルを更新することにより、高精度のパターン認識が可能なパターン認識辞書が高速に作成できる。

【0019】(6) 本発明のパターン認識方法(請求項6、固有ベクトルの学習:未定乗数法)で用いられる前記更新処理は、前記パターンベクトルと予め定められた制限のある辞書データ(正規直行ベクトル ψ_i)との類似度計算式(式(52))に従った認識結果に基づく第1の評価式(式(53))の第1項に、さらに前記辞書データの予め定められた制限を表した式に任意の定数を乗じたもの(式(53))の第2項を加算した第2の評

価式(式(53)あるいは式(54))を最適化する方向に前記辞書データを更新するための漸化式(式(57))に従ってカテゴリ毎の前記辞書データを更新することにより、高精度のパターン認識が可能なパターン認識辞書が高速に作成できる。

【0020】(7) 本発明のパターン認識方法(請求項7、ウインドウ関数)で用いられる前記更新処理は、前記パターンベクトルと辞書データ(例えば、 K)との類似度計算式 (s) に従った認識結果に基づく評価式

(J) を最適化する方向に前記辞書データを更新するための漸化式(例えば、式(13))に従ってカテゴリ毎の前記辞書データを更新し、前記漸化式には、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための関数で、前記正解カテゴリと不正解カテゴリの類似度の差が等しい部分で高い値を取り、そうでない部分で低い値をとる窓関数 $w(d)$ を有し、前記辞書データを更新する際、前記正解カテゴリと不正解カテゴリの類似度の差が所定の範囲内の時のみに前記窓関数を用いることにより、高精度のパターン認識が可能なパターン認識辞書が高速に作成できる。

【0021】(8) 本発明のパターン認識方法(請求項8、ウオッチドッグ方式、図4参照)で用いられる前記更新処理は、前記パターンベクトルと辞書データとの類似度計算式に従った認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従ってカテゴリ毎の前記辞書データを更新し、前記漸化式には、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための窓関数を有し、この窓関数の特性を定めるパラメータ値は、前記パターンベクトルの複数の認識結果を比較することにより設定される。すなわち、パターン認識処理または別のテストベクトルによる認識テストの結果をチェックし、認識性能が向上しないか、または悪くなる状況に相当する場合にはパラメータの値を設定しなおして、必要な情報を復帰させて学習を再スタートさせるようにし、この動作を繰り返すことにより、高精度のパターン認識が可能なパターン認識辞書が高速に作成できる。

【0022】(9) 本発明のパターン認識方法(請求項9、アダプティブウインドウ関数を用いるもの、図6参照)で用いられる前記更新処理は、前記パターンベクトルと辞書データとの類似度計算式に従った認識結果に基づく評価式を最適化する方向に前記辞書データを更新し、前記漸化式には、少なくとも認識結果として得られる正解カテゴリと不正解カテゴリの類似度によって定まる重み値を決定するための窓関数を有し、この窓関数の特性を定めるパラメータ値は、前記パターンベクトルの認識結果に基づく前記正解カテゴリと不正解カテゴ

リの類似度に関する頻度分布に基づき設定されることにより、高精度のパターン認識が可能なパターン認識辞書が高速に作成できる。

【0023】(10) 本発明のパターン認識方法(請求項10、大分類対応)で用いられる前記更新処理は、前記パターンベクトルと辞書データとの類似度計算式に従った該パターンベクトルの大分類の認識結果に基づく評価式を最適化する方向に前記辞書データを更新するための漸化式に従って正解カテゴリと不正解カテゴリの前記辞書データを更新することにより、高精度の大分類パターン認識が可能な大分類パターン認識辞書が高速に作成できる。

【0024】

【発明の実施の形態】以下、本発明の実施形態について図面を参照して説明する。

【0025】(第1の実施形態)図1は、本発明の第1の実施形態に係るパターン認識辞書作成装置の構成例を示したもので、学習パターンを入力するためのパターン入力部1、認識処理部2、辞書更新部3、初期辞書作成部4、辞書データ格納部5から構成される。

【0026】まず、初期辞書作成部4では、パターン入力部1から入力された学習パターンから初期辞書を作成して、その作成された初期辞書を辞書データ格納部5に格納すると、次に、認識処理部2では、入力された学習*

$$l(x) = 1/(1 + e^{-x})$$

【0031】とする。このとき、入力ベクトル x の認識結果に対する評価を行うための評価式を ※

$$J = \int_{x \in \Omega} l(s_{er}(x) - s_{ok}(x)) p(x) dx. \quad \dots (3)$$

$$s_{ok}(x) = \max_{C(x)=c} x^T K_c x = x^T K_{ok} x \quad \dots (4)$$

$$s_{er}(x) = \max_{C(x) \neq c} x^T K_c x = x^T K_{er} x. \quad \dots (5)$$

【0032】とする。ここで、 s_{ok} は、 x を各変換行列 K に基づく類似度計算式(1)により認識した時の、認識正解カテゴリに対応する変換行列に対する類似度に対応しており、 s_{er} は不正解カテゴリに対応する類似度である。このケースではそれぞれ最大値を示すカテゴリとしている。

【0033】なお、正解カテゴリとは認識結果として正解となるという意味ではなく、入力データの答えとなるべきカテゴリという意味である。不正解カテゴリについても同様である。

【0034】以下、説明を分かり易くするために損失関

* パターンに対し、辞書データ格納部5に格納されている初期辞書または学習途中の辞書を用いて認識処理を行い、辞書更新部3で、その結果と入力された学習パターンに基づいて辞書データ格納部5に格納された認識辞書を更新するようになっている。この更新を終了条件が満たされるまで繰り返すことにより辞書の学習、改良を行って、パターン認識辞書を作成する。

【0027】以下、主に、認識処理部2と辞書更新部3の処理動作について説明する。

【0028】1) 特性核学習

まず、本発明のパターン認識辞書作成方法に用いられる辞書更新方法(特性核学習)について説明する。

【0029】入力ベクトルを x とし、それぞれのカテゴリに対して K を x に対する変換行列、類似度 s を

$$s = x^T K x \quad \dots (1)$$

と定義する。入力ベクトル x について各カテゴリ中で最大の類似度を持つカテゴリを x の認識結果として出力するものとする。

【0030】 Ω を全学習パターンセット、 $C(x)$ を x のカテゴリ名、 K_c をカテゴリ名 c に対応する変換行列、 M を全パターン数、 $p(x)$ を x の出現確率、 $l(x)$ を損失関数とする。ここで、損失関数 $l(x)$ は、

$$\text{【数1】} \quad \dots (2)$$

※ 【数2】

数1を

$$l(s_{er}(x_k) - s_{ok}(x_k))$$

と表すが、 $l(s_{er}(x_k), s_{ok}(x_k))$ とで定義しても同様な考え方で学習方式を設定することができる。

【0035】式(4)の K_{ok} は、正解カテゴリに対応する変換行列、 K_{er} は、不正解カテゴリに対応する変換行列である。

【0036】次に、評価式(3)の値 J を最小化する場合を考える。式(3)を離散的に表すと、

【数3】

(8)

特開平 11-154146

14

$$J = \sum_{k=1}^{13} l(s_{er}(x_k) - s_{ok}(x_k)).$$

... (6)

【0037】となる。これをKの要素 k_{ij} で微分する * 【数4】
と、

$$\frac{dJ}{dk_{ij}} = \sum_{k=1}^M l'(s_{er}(x_k) - s_{ok}(x_k)) \left(\frac{ds_{er}}{dk_{ij}} - \frac{ds_{ok}}{dk_{ij}} \right). \quad \dots (7)$$

ここで、

$$\omega(x) = l'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = l(x)(1-l(x)). \quad \dots (8)$$

【0038】また、 $d_k = s_{er}(x_k) - s_{ok}(x_k)$ ※ (9) のように表せるので、
と置いて、 $w(d) = l'(d)$ と表す。 $w(d)$ は対 【数5】
照関数である。Nを次元数とし、また、類似度sは式※

$$s = \sum_{i=1}^N \sum_{j=1}^N x_i x_j k_{ij} \quad \dots (9)$$

$$(k_{ij}) = K_{ok} \text{ の時、 } \frac{ds_{ok}}{dk_{ij}} = x_i x_j, \quad \frac{ds_{er}}{dk_{ij}} = 0. \quad \dots (10)$$

$$(k_{ij}) = K_{er} \text{ の時、 } \frac{ds_{ok}}{dk_{ij}} = 0, \quad \frac{ds_{er}}{dk_{ij}} = x_i x_j. \quad \dots (11)$$

【0039】となる。最急降下法によれば、Jを最小に ★ 【数6】
するためには、

$$k'_{ij} = k_{ij} - \alpha \frac{dJ}{dk_{ij}} = k_{ij} \pm \alpha \omega(d) x_i x_j, \quad \begin{cases} (k_{ij}) = K_{ok} (ok) \\ (k_{ij}) = K_{er} (error) \end{cases} \quad \dots (12)$$

【0040】とすれば良い。すなわち、Kを更新するた 40 ☆ 【0041】
めの漸化式(学習漸化式)は、次式(13)のように表 【数7】
すことができる。 ☆

$$K' = K \pm \alpha \omega(d) x x^T, \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (13)$$

【0042】なお、行列Kは、辞書データ格納部5に格納される辞書データであり、各カテゴリに対応して行列Kが存在し、この行列Kを更新することにより、辞書を作成するようになっている。

【0043】次に、図3に示すフローチャートを参照し 50

て、図1のパターン認識辞書作成装置の処理動作を説明する。

【0044】まず、初期行列 K_0 を作成する(後述)。各カテゴリを示す変数iを「0」に初期化し、以下の処理を繰り返す。

【0045】ステップS3：学習用の入力ベクトル x が入力されると、辞書データ格納部5に各カテゴリ毎の行列 K_i を用いて類似度計算を行う。

【0046】ステップS4：最大類似度を示す正解カテゴリの行列 K_i に対しては、式(13)に示すように、 $\alpha \omega(d) x x^T$ を加算し、最大類似度を示す不正解カテゴリの行列 K_i に対しては、同じく式(13)に示すように、 $\alpha \omega(d) x x^T$ を減算して、行列 K_i を更新する。

【0047】ステップS5～ステップS6： i を「1」ずつインクリメントし、所定の終了条件を満たすまで、全てのカテゴリについて上記ステップS3～ステップS4の処理を実行する。

【0048】終了条件は、たとえば、ある回数で終了させるようにしても良いし、認識率などの数値データに基づいて停止させるようにしても良い。

【0049】2) 特性核学習の第2の例
辞書を更新する際、すなわち、式(13)を用いて K から K' を作成する際、公知例として示したALSMと同様に、正解の時はなにもせず、不正解の時に、
不正解カテゴリに対して $K' = K + \alpha x x^T$ *

$$K' = K \mp \alpha \omega(d)(x - m)(x - m)^T.$$

m は、

$$m' = m \pm 2\alpha \omega(d)K(x - m). \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (15)$$

で更新する。

【0053】この結果を用いて認識する際には、行列 K から得られる固有ベクトルの一部を用いた類似度式を構成して、これにより認識を行うように構成することもできる。

【0054】3) 初期行列 K_0 の作成方法
図3のステップS1の初期行列 K_0 は以下のようにして ※

$$Q = \sum_{x \in \Omega_{all}} x x^T,$$

【0056】なる共分散行列 Q を作成し、その固有ベクトルを求める。そして、各固有ベクトルに対応する固有値を比較して、その値の大きい固有ベクトルからなる行列 T を作る。 ★

$$Q' = \sum_{x \in \Omega'} (Tx)(Tx)^T,$$

【0058】なる共分散行列 Q' を作成し、その固有ベクトル、固有値を求める。そして、固有ベクトルからなる直交行列 U 、固有値を対角に並べて他を「0」とした対角行列 Λ を作り、 $Q' = U^{-1} \Lambda U$ と表すとす。こ

* 不正解カテゴリに対して $K' = K - \beta x x^T$ として、 K を更新することもできる。ここで、 α 、 β は任意の定数である。

【0050】この方法の従来例はさきに述べたALSMであるが、ALSMでは類似度 s の計算式が $s = x^T P x$ であるのに対して、本発明の学習漸化式を導き出すために用いた評価式(6)で用いる類似度 s が $s = x^T K x$ とく違っており、理論的に誤った学習法を提供するものであった。本発明では、上記の説明でわかる通り正しい類似度計算と学習漸化式を提供している。これにより効果的に性能の良い認識系を構成することができる。

【0051】さて、前述した類似度計算式(1)や学習漸化式(13)に平均ベクトルに相当するベクトル m を導入して、相違度を

$$s = (x - m)^T K (x - m) \quad \dots (14)$$

と定義することもできる。この場合の学習漸化式は、式(13)の1の中の s_{er} と s_{ok} が入れ替わることを考慮して、次式(15)のように表すことができる。

【0052】

【数8】

$$\begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (15)$$

※求めることができる。

【0055】まず、全学習ベクトル(上記逐次学習に用いたものでも、また違うものでも良い)の集合を Ω'

とし、

【数9】

$$\dots (17)$$

★【0057】次に、各カテゴリ毎の学習ベクトルの集合を Ω' とし、

【数10】

$$\dots (18)$$

の直行列 U を構成する固有ベクトルの中で、対応する固有値の大きいものからなる行列 P を作り、初期変換行列 $K_0 = T^T P^T P T = (P T)^T (P T)$ を作成する。

【0059】上記の手法は T の求め方としてパターン認

識分野で良く知られたKL変換を用いたものであるが、
そのかわりに、やはりパターン認識分野で良く知られて
いる正準判別を用いても良い。即ち、級間共分散行列 *

* S_b 、級内共分散行列 S_w を用いて、
【数11】

$$S_b \varphi = \lambda S_w \varphi,$$

… (19)

【0060】を満たす固有ベクトルからTを作成する方
法である。

※は正値対称行列である。以下、 $K^{-1} = H$ と表記する。こ
の場合、評価式(3)は、離散的に表すと、式(20)
に示すようになる。

【0061】4) 特性核学習の第3の例

ここでは、類似度として重み付き部分空間法を用いる場
合について説明する。

【0063】

10 【数12】

【0062】相違度 s を $s = x^T K^{-1} x$ と定義する。K ※

$$J = \sum_{k=1}^M l(s_{ok}(x_k) - s_{er}(x_k)).$$

… (20)

【0064】なお、評価式(3)では類似度であった
が、ここでは相違度なので、式(20)において、 s_{ok}
(x_k)と $s_{er}(x_k)$ の位置が入れ替わっている。式

★得られる。

【0065】

【数13】

(20)をHの要素 h_{ij} で微分すると、次式(21)が★

$$\frac{dJ}{dh_{ij}} = \sum_{k=1}^M l'(s_{ok}(x_k) - s_{er}(x_k)) \left(\frac{ds_{ok}}{dh_{ij}} - \frac{ds_{er}}{dh_{ij}} \right).$$

… (21)

ここで、 $\omega(d) = l'(s_{ok}(x_k) - s_{er}(x_k))$ と表す。

【0066】また、相違度 s は

☆ ☆ 【数14】

$$s = \sum_{i=1}^N \sum_{j=1}^N x_i x_j h_{ij}$$

… (22)

【0067】と表されるので、

◆ ◆ 【数15】

$$(h_{ij}) = H_{ok} \text{ の時、 } \frac{ds_{ok}}{dh_{ij}} = x_i x_j, \quad \frac{ds_{er}}{dh_{ij}} = 0.$$

… (23)

$$(h_{ij}) = H_{er} \text{ の時、 } \frac{ds_{ok}}{dh_{ij}} = 0, \quad \frac{ds_{er}}{dh_{ij}} = x_i x_j.$$

… (24)

【0068】が得られる。最急降下法によれば、Jを最
小にするためには、

* 【数16】

$$h'_{ij} = h_{ij} - \alpha \frac{dJ}{dh_{ij}} = h_{ij} \mp \alpha \omega(d) x_i x_j.$$

*

$$\begin{cases} (h_{ij}) = H_{ok} \\ (h_{ij}) = H_{er} \end{cases}$$

… (25)

【0069】とすれば良い。すなわち、Hを更新するた
めの漸化式は、次式(26)のように表すことがで
きる。

※ 【0070】

【数17】

※

$$H' = H \mp \alpha \omega(d) x x^T, \quad \begin{cases} H = H_{ok} \\ H = H_{er} \end{cases}$$

… (26)

【0071】ここで、 $H' = H + \delta H$ とし、 $(H')^{-1}$
 $= K + \delta K$ と置くと、

【数18】

$$I = (H + \delta H)(H + \delta H)^{-1} = (H + \delta H)(K + \delta H) =$$

$$I + \delta H \cdot K + H \cdot \delta K + \delta H \cdot \delta K \approx I + \delta H \cdot K + H \cdot \delta K$$

$$\delta H \cdot K = -H \cdot \delta K \quad K \cdot \delta H \cdot K = K \cdot H \cdot \delta K = -\delta K, \quad \delta K = -K \cdot \delta H \cdot K$$

【0072】一方、 * * 【数19】

$$\delta H = \mp \alpha \omega(d) x x^T$$

であるから

$$\delta K = \pm \alpha \omega(d) K x x^T K$$

$K^T = K$ なので

$$\delta K = \pm \alpha \omega(d) K x (K x)^T$$

K を更新する漸化式は

$$K' = K \pm \alpha \omega(d) K x (K x)^T \quad \dots (27)$$

となる。

【0073】なお、相違度計算として次元打ち切りを行 ※ 【0074】

なって、実際には以下のような変換により類似度として 【数20】

計算するのが球面疑似ベイズの考え方である。

$$s = \sum_{i=1}^N \frac{1}{\lambda_i} (x, \varphi_i)^2 \approx \sum_{i=1}^r \frac{1}{\lambda_i} (x, \varphi_i)^2 + \sum_{i=r+1}^N \frac{1}{h} (x, \varphi_i)^2 = \sum_{i=1}^r \left(\frac{1}{\lambda_i} - \frac{1}{h} \right) (x, \varphi_i)^2 + \frac{1}{h} \|x\|^2$$

$$= \frac{1}{h} \left[\|x\|^2 - \sum_{i=1}^r \left(1 - \frac{h}{\lambda_i} \right) (x, \varphi_i)^2 \right] \quad \dots (28)$$

【0075】この場合は $\|x\|^2$ が一定なので、式 (2) ★いて説明する。

8) の第2項を類似度として計算すればよい。

【0077】相違度を

【0076】5) 特性核学習の第4の例

【数21】

ここでは、ベイズ識別に対応する方法を用いた場合につ★40

$$s = (x - m)^T K^{-1} (x - m) \quad \dots (29)$$

【0078】と定義し、第3の辞書更新方法の説明と同
様にして、 K を更新するための学習漸化式を導くと、次
式 (30) が得られる。 ☆

☆ 【0079】

【数22】

$$K' = K \pm \alpha \omega(d) K (x - m) (K (x - m))^T \quad \dots (30)$$

【0080】一方、 m については次の様にして m を更
新するための漸化式を求める。すなわち、評価式は

【数23】

(12)

$$J = \sum_{k=1}^M I(s_{ok}(x_k) - s_{er}(x_k)).$$

特開平 1 1 - 1 5 4 1 4 6

22

... (3 1)

$$s = \sum_{i=1}^N \sum_{j=1}^N (x_i - m_i)(x_j - m_j)h_{ij}.$$

... (3 2)

$$\frac{dJ}{dm_p} = \pm l'(d) \left\{ \sum_j (-1)(x_j - m_j)h_{pj} + \sum_i (x_i - m_i)(-1)h_{ip} \right\}$$

$$= \mp 2l'(d) \sum_i h_{pi}(x_i - m_i) \quad \begin{cases} ok. \\ error. \end{cases}$$

... (3 3)

ここで $h_{ij} = h_{ji}$ に注意。

$$\delta m = \pm 2l'(d)H(x - m).$$

$$m = m \pm 2\alpha\omega(d)H(x - m).$$

【0081】疑似ベイズでは、打ち切り計算を次のよう
に行なう。

30

【0082】

【数24】

$$H = \sum_{i=1}^N \frac{1}{\lambda_i} \varphi_i \varphi_i^T$$

であるから、

$$\begin{aligned} H(x-m) &= \sum_{i=1}^N \frac{1}{\lambda_i} (x-m, \varphi_i) \varphi_i \approx \sum_{i=1}^r \frac{1}{\lambda_i} (x-m, \varphi_i) \varphi_i + \sum_{i=r+1}^N \frac{1}{h} (x-m, \varphi_i) \varphi_i \\ &= \frac{1}{h} \left\{ \sum_{i=1}^r \left(\frac{h}{\lambda_i} - 1 \right) (x-m, \varphi_i) \varphi_i + (x-m) \right\}. \end{aligned} \quad \dots (35)$$

よって、漸化式は以下のようになる。

$$m = m \pm \frac{2\alpha\omega(d)}{h} \left\{ (x-m) - \sum_{i=1}^h \left(1 - \frac{h}{\lambda_i} \right) (x-m, \varphi_i) \varphi_i \right\}. \quad \dots (36)$$

【0083】6) ALSMを用いた特性核学習：ウインド関数を用いる場合
公知例として示したALSMでは、学習漸化式のなかでは損失関数微分は使われておらず、定数値の α 、 β が使われていたが、これは、学習を的確に行うには不適當であった。学習結果が正解と不正解のボーダーラインでの*

* 確に行われるように損失関数の微分に相当する窓関数を用いることが望ましい。そこで、本発明では、正解カテゴリに対する類似度と不正解カテゴリに対する類似度が等しい部分で、そうで無い時よりも高い値を取る窓関数 $w(d)$ を用いて学習漸化式を

【数25】

$$K' = K \pm \alpha \omega(d) x x^T, \quad \begin{cases} K = K_{\text{ok}} \\ K = K_{\text{err}} \end{cases} \quad \dots (37)$$

【0084】で構成している。なお、窓関数としては、例えば式(7)を用いることができる。

* はKの固有ベクトルからなる行列である。この類似度をKの固有ベクトル ψ を用いて書き換えると

【数26】

【0085】ここでの類似度は $s = x^T P x$ である。P *

$$s = \sum_{i=1}^r (x, \varphi_i)^2 \quad (r < N) \quad \dots (38)$$

【0086】となる。

★面ベイズと呼ばれる、

【数27】

【0087】この類似度計算式(38)には、上記のALSMタイプ以外の方法も有効である。例えば、疑似球★40

$$s = (x, \varphi)^2 + \sum_{i=1}^r \left(1 - \frac{\delta}{\lambda_i} \right) (x, \varphi_i)^2 + c. \quad \dots (39)$$

【0088】がある。ここで、 λ_i はKの固有値、 δ 、 c は固有値から決まるスカラー値である。

平均ベクトルとを更新するALSMを用いた特性核学習について説明する。

【0089】7) ALSMを用いた特性核学習に定数、平均ベクトルの学習を導入

【0090】前述した類似度計算式や漸化式に平均ベクトルに相当するベクトル m を導入して、相違度を投影距離、すなわち、

【数28】

また、前述のALSMを用いた特性核学習に、さらに平均ベクトルを導入する場合を考える。すなわち、辞書データ格納部5に格納される辞書データとしての行列Kと 50

$$s = \|x - m\|^2 - \sum_{i=1}^r (x - m, \varphi_i)^2 \quad (r < N),$$

… (40)

【0091】と定義することもできる。この場合の行列 K の漸化式は、

$$K' = K \mp \alpha \omega(d) (x - m)(x - m)^T, \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (41)$$

【0092】であり、 m は、実際に平均ベクトルを学習ベクトルから計算して求めても良いが、以下のように漸化式で更新することもできる。

$$m' = m \pm 2\alpha \omega(d) K (x - m), \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (42)$$

【0094】この場合に疑似ベイズタイプの相違度、すなわち

$$s = \|x - m\|^2 - \sum_{i=1}^r (1 - \frac{\delta}{\lambda_i}) (x - m, \varphi_i)^2 + c \quad (r < N), \quad \dots (43)$$

【0095】も有効である。

20 ☆ 【数32】

【0096】式 (43) に含まれる定数 c は、

☆

$$c' = c \mp \alpha \omega(d), \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (44)$$

【0097】によって更新することもできる。

【0098】8) 固有ベクトルの学習

(1) 2次形式評価式タイプで係数、定数学習を含むもの (ノルムが入っていてもよい)

◆ 次に、 K の固有ベクトル ψ を直接逐次学習して辞書更新を行う場合について説明する。すなわち、例えば、類似度を、

【数33】

$$s = \sum_{i=0}^r a_i (x, \varphi_i)^2 + c \quad \dots (45)$$

$$s = \sum_{i=0}^r a_i (x, \varphi_i)^2 + \sum_{i=0}^r b_i (x, \varphi_i) + c \quad \dots (46)$$

$$s = \|x - m\|^2 - \sum_{i=0}^r a_i (x - m, \varphi_i)^2 + \sum_{i=0}^r b_i (x - m, \varphi_i) + c \quad \dots (47)$$

【0099】のうちのいずれかで定義し、これに基づいた評価式を

* 【数34】

$$J = \sum_{x \in \Omega} l(s_{er}(x) - s_{ok}(x)) \quad \dots (48)$$

【0100】で設定し、この評価式を最適化する方向に、逐次、 ψ 、 m 、 a_i 、 b_i を更新することにより、より良い類似度や相違度計算のための辞書データを決定する方法である。

【0101】従来は、 ψ 、 m のみを学習していたので、本実施形態により作成された認識辞書を用いればパター

ン認識能力は従来例よりも精度が高くなる。

【0102】逐次変更する手段としては、たとえば、従来の L S M と同じ手段で ψ_i や m を更新すれば良い。すなわち、

【数35】

$$\varphi'_i = \varphi_i + \alpha(R(x))(x, \varphi_i)x$$

28
... (49)

$$m' = m + 2\alpha(R(x))(x - m) - \sum_{i=0}^r (\varphi_i, x - m)\varphi_i \quad \dots (50)$$

【0103】 ψ_i の正規直交性の保証は従来例のようにシュミットの直交化を用いてもいいし、また、後述の第4の実施形態に示す方法を用いても良いし、その他の方法も可能である。

【0104】従来例に無かった a_i 、 b_i 、 c については、類似度を例えば式(47)で定義するとすれば、次の様に更新する。

【0105】

【数36】

$$a'_i = a_i + \alpha(R(x))(x - m, \varphi_i)^2,$$

$$b'_i = b_i + \alpha(R(x))(x - m, \varphi_i),$$

20

$$J = \sum_{x \in \Omega} \left\{ l(s_{er}(x) - s_{ok}(x)) + A \sum_i \sum_j (\delta_{ij} - (\varphi_i, \varphi_j))^2 \right\} \quad \dots (51)$$

$$\delta_{ij} = \begin{cases} 1(i=j), \\ 0(i \neq j), \end{cases} \quad s_{ok}(x) = \max_{C(x)=c} s_c(x) = s_{ok}(x) \quad s_{er}(x) = \max_{C(x) \neq c} s_c(x) = s_{er}(x).$$

【0108】この評価式を最適化する方向に各パラメータ、ベクトルを学習していくことにより学習を行う。以下、具体的に説明する。

※【0109】まず、類似度の定義を次のように行なう。
【数38】

$$s = \sum_{i=0}^r a_i(x, \varphi_i)^2 + c. \quad \dots (52)$$

【0110】評価式は、
【数39】

$$(\varphi_i, \varphi_j) = \begin{cases} 1(i=j), \\ 0(i \neq j). \end{cases}$$

40★【0111】である事を考慮し、正則化の考え方と同様な手法で係数Aを導入して
【数40】

$$J = \sum_{x \in \Omega} \left\{ l(s_{er}(x) - s_{ok}(x)) + A \sum_i \sum_j (\delta_{ij} - (\varphi_i, \varphi_j))^2 \right\} \quad \dots (53)$$

【0112】と評価式を設定する。

【0113】式(53)の値Jを最小化することを考え

る。式(53)を離散的に表すと、
【数41】

$$J = \sum_{k=1}^M l(s_{er}(x_k) - s_{ok}(x_k)) + A \sum_i \sum_j (\delta_{ij} - (\varphi_i, \varphi_j))^2$$

… (54)

となる。ここで s は、

$$s = \sum_{i=0}^r a_i \left(\sum_{j=0}^n x_j \varphi_{ij} \right)^2 + c$$

であり、

$$X = A \sum_i \sum_j (\delta_{ij} - (\varphi_i, \varphi_j))^2$$

と置く。

$$\frac{ds}{d\varphi_{pq}} = a_p \cdot 2(x, \varphi_p) \cdot x_q, \quad \frac{ds}{d\varphi_p} = 2a_p (x, \varphi_p) x.$$

【0114】ここで、もし、式(54)に X の項がない * 化式は
 とすると式(54)の右辺が漸化式の増分に相当し、漸* 【数42】
 $\varphi_p = \varphi_p \pm 2a_p \alpha \omega(d)(x, \varphi_p) x$ … (55)

【0115】となる。一方、

【数43】

$$\frac{ds}{da_p} = (\varphi_p, x)^2, \quad \frac{ds}{dc} = 1.$$

【0116】次に、 X について微分する。 X は正解カテゴリ、不正解カテゴリに対する類似度 s_{ok} 、 s_{er} に関係ない項である。

【0117】

【数44】

$$\frac{dX}{d\varphi_{pq}} = A \sum_i \sum_j 2(\delta_{ij} - (\varphi_i, \varphi_j))(\delta_{ij} - (\varphi_i, \varphi_j))' \quad \dots (56)$$

$(\delta_{ij} - (\varphi_i, \varphi_j))'$ については φ_{pq} で微分すると、

$$i = p \text{ の時に } -(\varphi_{pq})' \cdot \varphi_{jp} = -\varphi_{jp},$$

$$j = p \text{ の時に } -\varphi_{iq} \cdot (\varphi_{pq})' = -\varphi_{iq},$$

のみが残る。従って、

$$\begin{aligned} \frac{dX}{d\varphi_{pq}} &= A \sum_i \sum_j 2(\delta_{ij} - (\varphi_i, \varphi_j)) \{ -(\varphi_{pq})' \cdot \varphi_{jq} - \varphi_{iq} \cdot (\varphi_{pq})' \} \\ &= 2A \left\{ \sum_j (\delta_{pj} - (\varphi_p, \varphi_j)) \cdot (-\varphi_{jq}) + \sum_i (\delta_{ip} - (\varphi_i, \varphi_p)) \cdot (-\varphi_{iq}) \right\} \\ &= 2A \left\{ -\varphi_{pq} + \sum_j (\varphi_p, \varphi_j) \varphi_{jq} + -\varphi_{pq} + \sum_i (\varphi_i, \varphi_p) \varphi_{iq} \right\} \\ \frac{dX}{d\varphi_p} &= 2A - \varphi_p + \sum_j (\varphi_p, \varphi_j) \varphi_j + 2A - \varphi_p + \sum_i (\varphi_i, \varphi_p) \varphi_i \\ &= 4A \left\{ \sum_{i=0}^r (\varphi_i, \varphi_p) \varphi_i - \varphi_p \right\}. \end{aligned}$$

【0118】これらより ψ の漸化式は次の様に求まる。

【数45】

【0119】

$$\varphi_p = \varphi_p + \alpha \omega(d) \left[4A \left\{ \varphi_p - \sum_{i=0}^r (\varphi_i, \varphi_p) \varphi_i \right\} \pm 2a_p(x, \varphi_p)x \right] \begin{cases} ok. \\ error. \end{cases}$$

$$a_p = a_p \pm \alpha \omega(d) (\varphi_p, x)^2 \begin{cases} ok. \\ error. \end{cases}$$

$$c = c \pm \alpha \omega(d) \begin{cases} ok. \\ error. \end{cases} \quad \dots (57)$$

【0120】以上から部分空間法に変形する場合、
【数46】

$$a_i = 1, c = 1$$

【0121】であり、式(57)の漸化式がそのまま使える。

【0122】9) ウィンドウ関数

これまでに述べてきた学習において、漸化式で使われる窓関数 $w(d)$ に工夫を加えることによって更新処理をより高速化できる。

【0123】まず、 $w(d)$ と $d=0$ 付近で高い値を取り、周辺で低い値をとる関数(例えば、 $1'(x)$)を用意する。これに対してある閾値 TH を用意して
 $w(d) < TH$

の時は、学習しないように制御することにより、ほとんど役には立たないが、処理時間の要するものをカットするのである。この方法はさらに、

$$(d < TH1) \text{ or } (TH2 < d)$$

の時に学習しないように制御することもできる。もちろん、 $w(d)$ として $d=0$ 付近で高い値を取り、周辺で低い値をとり、 $TH1$ 、 $TH2$ の外側では完全にゼロとなる関数を用いて同じことである。

【0124】10) ウォッチドッグ方式

一般的に、学習時は、入力ベクトルを x 、辞書データを Λ とした時に、類似度を

$$s = S(x, \Lambda)$$

と定義して、認識結果として得られるデータ $R(x)$ から決められる窓関数 $w(R(x))$ と、入力ベクトルと辞書データなどから決められる学習データの $D(x, \Lambda)$ を用いて

$$\Lambda' = \Lambda + w(R(x)D(x, \Lambda))$$

と書ける。この時 w を適切に設定しないと学習がうまくいかないという問題点があり、この問題を解決する方法として、ウォッチドッグ方式を採用する。すなわち、ここでいうウォッチドッグ方式とは、窓関数のパラメータ*

$$\sigma(x) = \frac{1}{1 + e^{-\alpha x}},$$

* 値を設定する際、学習の際に行われる認識、またはテストデータによる認識テストの結果をチェックし、認識性能が向上しないか、または悪くなる場合に $w(R(x))$ の値が小さくなるように設定しなおして(その際、辞書データを等を設定し直して)、学習を再スタートさせるものである。図4に示すフローチャートを参照して、窓関数のパラメータの設定方法にウォッチドッグ方式を採用する場合のパターン認識辞書作成装置の処理動作について説明する。

【0125】まず、予め作成された辞書を用いて認識処理を行う(ステップS12)。次に認識結果に基づき学習を行う(ステップS13)。これをN回繰り返す(ステップS15～ステップS16)。

【0126】次に、認識結果を以前の認識結果と比較する。ただし比較が1回目の時は、認識精度が向上しているものとして以下を実行する。すなわち、ステップS17で認識処理結果を評価することにより、認識精度が以前のときより悪くなっているか、上がらない時は、ステップS19に進み、保存してあった辞書やパラメータなどを復帰させる。そして、 $w(R(x))$ をより小さな値になるように設定しなおして(ステップS19～ステップS20)、ステップS11に戻り、再度実行を行う(ステップS21)。

【0127】一方、ステップS17での評価の結果、認識精度が向上している時は、現在の辞書やパラメータを保存し(ステップS21)、ステップS11に戻り、実行を継続する。

【0128】これをM回繰り返しても改善が認められない時は学習を終了させる(ステップS21)。

【0129】ここで、窓関数として、次式(58)の微分を用いる場合を例にとり、より具体的にパラメータの設定方法について説明する。

【0130】

【数47】

$$\dots (58)$$

【0131】式(58)に示した関数 $\sigma(x)$ の概形、および、それを微分したもの $\sigma'(x)$ の概形を図5に

示す。図5には、式(58)の α の値が大きくなると、 $\sigma(x)$ の原点付近において、 x の値の変化に対する y の値の変化が大きくなり、そのカーブが急峻になる様子を(a)～(c)に示している。

【0132】損失関数としては、 α が大きい時の形、すなわちステップ型の方が適している。何故なら、正読の時はどの場合も損失は同じで、誤読の時もどの場合にも損失は同じと考えられるからである。しかしながら、完全なステップ関数では微分がゼロになってしまい学習しなくなってしまうので、無限大でない適当な α の値が窓関数としては必要となる。

【0133】そこで、学習の最初の時期には、 α の値を小さめに取って、徐々に大きくしていくことによって、より良い学習が行われるようになる。すなわち徐々に精密な学習をおこなわせていく考え方である。図4のステップS19において、 $w(R(x))$ の大きさを変更する際、式(58)の α の値を大きくするようにして学習を実行すれば良い。

【0134】11) アダプティブウィンドウ関数を用いるもの

一般的に言って、学習時の認識処理を行った時に、正解のカテゴリに対する類似度と不正解のカテゴリに対する類似度がきわどい時は、 $w(d)$ を大きめに取り、そうでない時は小さめにとることが望ましい。きわどい時は正解と不正解とを分ける境界線に対し、正解、不正解の2つパターンが近いところに存在するので、その結果を*

$$\omega(d) = \frac{e^{-\alpha d}}{(1 + e^{-\alpha d})^2}, \quad \alpha = D/v,$$

【0144】ここで、 D は、予め定められた適当な係数で、式(59)は、式(58)の微分形である。式(59)に示した窓関数は、図5に示したような山形をしていて、 v の値が小さくなると山の幅が小さくなるように変化する。したがって、 D を適切に選べば、ヒストグラムがどのような形をしていても、 $w(d)$ がちょうど良い幅をもつように設定できる。

【0145】学習の途中、たとえば N 回に1回、図4のステップS13における学習のための類似度計算の際に上記ヒストグラムを作成するようにし、そこで求めた v を使って次回以降の学習を行うようにする。

【0146】学習の過程で、分布の幅があまり変わらないようなケースでは最初に v の設定を行い、以降それに基づいた $w(d)$ によって学習を進めるような仕組みも有効である。

【0147】12) 大分類対応

一般に窓関数の値の決定や辞書の更新は最大類似度を取る正解カテゴリや不正解カテゴリについて行われるが、大分類時には別の方法が適している。大分類とは詳細認識に先立って行われるもので、認識候補を絞り込むためのものである。

* 使うことは妥当なのであるが、そうでない時は2つのパターンがかけ離れた場所にあるので境界線の学習に使うのは不適當であるからである。

【0135】そこで、以下のアルゴリズムによって、 $w(d)$ を設定する。

【0136】まず、認識結果から得られる類似度値の d を各学習パターンについて求め、これについてのヒストグラムを作成する。その一例を図6に示す。図6からまず、 $w(d)$ の幅に相当するパラメータ v を決める。その決め方としては、例えば、以下のものが挙げられる。

【0137】・分布のピーク値 P の半分の値を T とし、その位置に相当する横軸上の値と「0」との差の絶対値を v とする。

【0138】・ $d=0$ から右側の分布部分について、横方向に存在する位置の平均を計算し、それを v とする。

【0139】・ $d=0$ から右側の全分布面積に対して、 $d=0$ の位置から v の位置までの面積が $\alpha (<1)$ 倍の面積になる位置を v とする。

【0140】・ $d=0$ における微分係数に相当する数値から v を定める。

【0141】・上記の組み合わせで v を定める。

【0142】次に、この v を用いて、窓関数 $w(d)$ を例えば次式(59)によって定義する。

【0143】

【数48】

$$\dots (59)$$

【0148】以下、認識候補を p 個に絞り込む場合を例にとり説明する。この場合、正解カテゴリが p 位までに入っていれば良いので、次のように学習を行わせる。

【0149】正解カテゴリが p 位までに入っている時は $p+1$ 位のカテゴリの類似度を s_{er} 、辞書を Λ_{er} とし、そうで無いときは、 p 位のカテゴリの類似度を s_{er} 、辞書を Λ_{er} とする。一方、正解カテゴリの類似度を s_{α} 、辞書を Λ_{α} とする。窓関数を $w(s_{er}, s_{\alpha})$ として、入力ベクトルと辞書データなどから求められる学習データの $D(x, \Lambda)$ を用いて、 Λ_{er} 、 Λ_{α} について、漸化式 $\Lambda' = \Lambda + w(R(x)) D(x, \Lambda)$ によって辞書の更新を行う。

【0150】上記の例で、正解カテゴリが大分類候補に入っている場合には辞書の更新を行わないように制御しても良いし、窓関数として一定値を使うようにしても良い。また、 p 位以内のすべての不正解カテゴリについて辞書更新を行うように制御しても良いし、正解カテゴリが大分類候補に入っていない場合に、正解カテゴリより上位の不正解カテゴリについて辞書の更新を行うようにしても良い。たとえば、この場合、不正解カテゴリの類似度を s_{er} 、辞書を Λ_{er} として上記と同様にして辞書の

更新を行う。

【0151】（第2の実施形態）図2は、本発明の第2の実施形態に係るパターン認識装置の構成例を示したもので、前述したパターン認識辞書作成方法およびそれを用いた図1に示したようなパターン認識辞書作成装置を用いて作成された辞書を用いて入力パターンの認識を行うものである。辞書データ格納部5には、前述の辞書作成方法にて作成された辞書が格納され、認識処理部2では、パターン入力部1を介して入力されるパターンベクトルに対し、辞書データ格納部5に格納された辞書データを参照して、パターン認識を行い（入力パターンベクトルと辞書データと類似度 s あるいは相類似度 s を算出）、類似度が最も大きいカテゴリ、あるいは、相類似度が最も小さいカテゴリを認識結果出力部6から出力するようになっている。

【0152】なお、上記第1～第2の実施形態に記載した図1および図2の各構成部の処理は、コンピュータに実行させることのできるプログラムとして磁気ディスク（フロッピーディスク、ハードディスクなど）、光ディスク（CD-ROM、DVDなど）、半導体メモリなどの記録媒体に格納して頒布することもできる。

【0153】また、本発明は上記実施形態に限定されるものではなく、その要旨を逸脱しない範囲で種々変形して用いることができる。

【0154】

【発明の効果】以上説明したように、本発明によれば、認識結果に応じて適切な方向に認識辞書が更新されていくことが理論的に説明できる方式を用いているので、認識辞書は学習用パターンデータに対して認識能力が逐次改善されていく方向に変化していく。従って、高精度の

*【0155】すなわち、これまで理論的に不整合のあった種々の辞書学習方式に対して理論的に整合のとれた学習方式が提供されるので理論的に適切な最適な辞書作成が行え、結果的に高精度なパターン認識装置を実現することができる。

【0156】また学習を高速におこなう工夫（例えば、従来方式であるGPDのような行列Pと行列Qとの変換を行わない）によって実用的な時間で辞書作成ができる点でも効果的である。

10 【図面の簡単な説明】

【図1】本発明の第1の実施形態に係るパターン認識辞書作成装置の構成例を概略的に示した図。

【図2】本発明の第2の実施形態に係るパターン認識装置の構成例を概略的に示した図。

【図3】図1のパターン認識辞書作成装置の処理動作を説明するためのフローチャート。

【図4】窓関数のパラメータの設定方法にウオッチドッグ方式を採用する場合のパターン認識辞書作成装置の処理動作について説明するためのフローチャート。

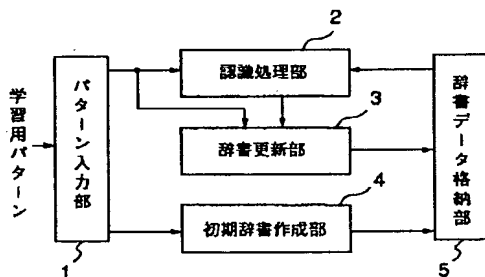
20 【図5】パラメータ値の変化に伴う窓関数の概形の変化を示した図。

【図6】パターンベクトルの認識結果に基づく正解カテゴリと不正解カテゴリの類似度の差の頻度分布の一例を示した図。

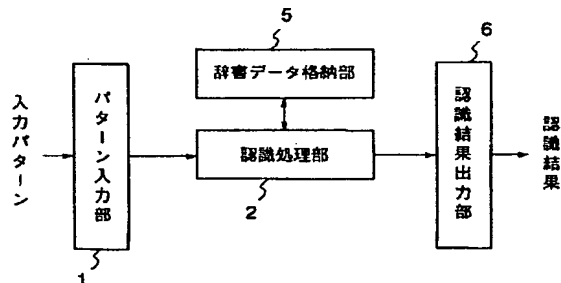
【符号の説明】

- 1…パターン入力部
- 2…認識処理部
- 3…辞書更新部
- 4…初期辞書作成部
- 5…辞書データ格納部
- 6…認識結果出力部

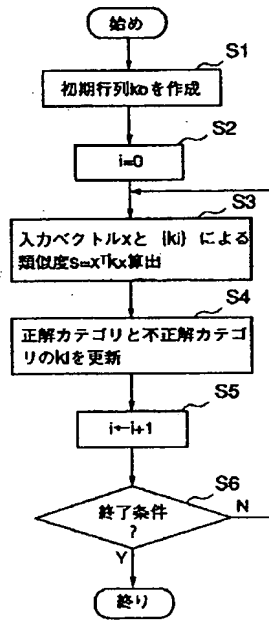
【図1】



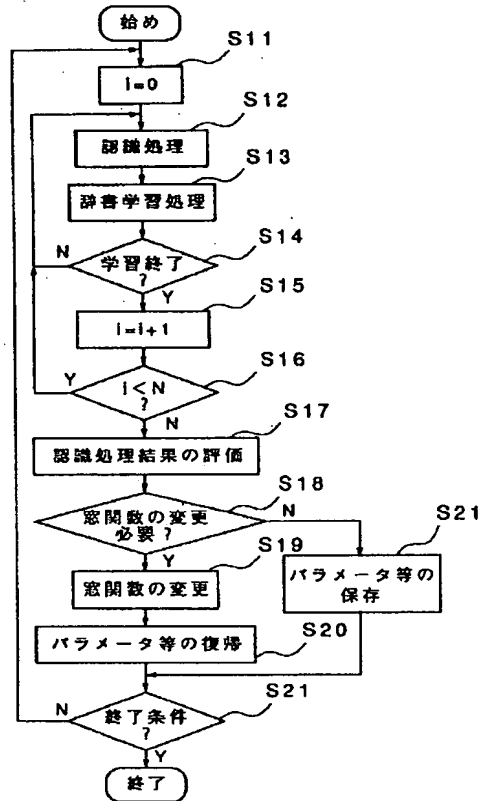
【図2】



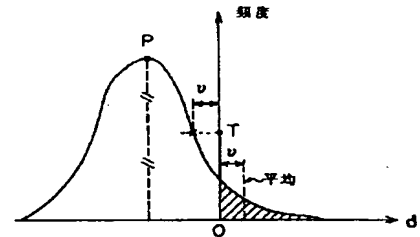
【図3】



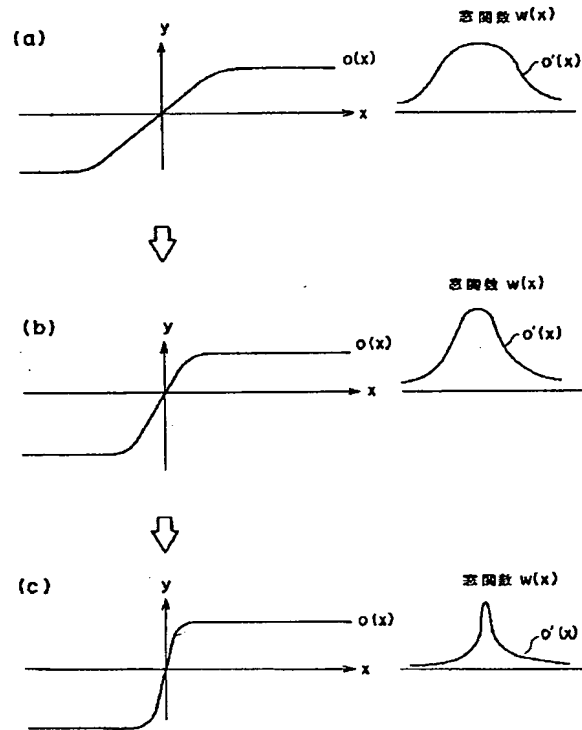
【図4】



【図6】



【図 5】



PATENT ABSTRACTS OF JAPAN

(11)Publication number : 11-154146

(43)Date of publication of application : 08.06.1999

(51)Int.Cl. G06F 15/18

G10L 3/00

G10L 3/00

// G06T 7/00

(21)Application number : 09-322883 (71)Applicant : TOSHIBA CORP

(22)Date of filing : 25.11.1997 (72)Inventor : KUROSAWA YOSHIKI

(54) PATTERN RECOGNITION METHOD AND RECORDING MEDIUM

(57)Abstract:

PROBLEM TO BE SOLVED: To theoretically balance a whole system, to execute learning at high speed and to generate a highly precise dictionary by updating matrixes for respective categories in accordance with a recurrence formula updating the matrixes in a direction for optimizing a value based on a recognition result.

SOLUTION: A recognition processing part 2 executes a recognition processing on an inputted learning pattern by using an initial dictionary stored in a dictionary data storage part 5 or a dictionary in the middle of learning. A dictionary update part 3 updates a recognition dictionary stored in the dictionary data storage part 5 based on the result and the inputted learning pattern. The matrixes are stored in the dictionary data storage part 5, the matrixes exist in accordance with the categories and the matrixes are updated accordance with the recurrence formula for updating the matrixes. Thus, the dictionary is generated. Namely, the recognition dictionary is updated in the appropriate direction in accordance with the recognition result and therefore recognition ability against the learning pattern is successively improved in the recognition dictionary.

LEGAL STATUS [Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

*** NOTICES ***

JP0 and NCIPi are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.

2.**** shows the word which can not be translated.

3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] In the pattern recognition approach of recognizing the inputted pattern using the pattern recognition dictionary drawn up by repeating an update process of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector Said update process imposes a matrix on said pattern vector. The recurrence formula which updates said matrix in the direction which optimizes the value of a valuation plan based on the recognition result of having followed the similarity formula for computing the similarity of this vector to each category containing the formula which furthermore hangs transposition of this pattern vector is followed. The pattern recognition approach characterized by updating said matrix for every category.

[Claim 2] The initial matrix of the matrix for said every category is the pattern recognition approach according to claim 1 characterized by being the matrix computed from the 1st matrix computed based on the set of said pattern vector, and the 2nd matrix for every category computed based on the set of said pattern vector classified into each category.

[Claim 3] In the pattern recognition approach of recognizing the inputted pattern using the pattern recognition dictionary drawn up by repeating an update process of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector Said update process imposes the 1st matrix on said pattern vector. Furthermore, transposition of this pattern vector The recurrence formula which updates said 1st matrix in the direction which optimizes the value of a valuation plan based on the recognition result of having followed the similarity formula for calculating the similarity of this pattern vector to each category based on the 2nd matrix which consists of a characteristic vector of the 1st formula to hang and said 1st matrix is followed. Said 1st matrix containing said characteristic vector for every category is updated. To said recurrence formula With the function for determining the weight value which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category The pattern recognition approach characterized by having the windowing function which takes a low value in the part from which the difference of the similarity of said correct answer category and unjust solution category takes a high value in an equal part, and the difference of the similarity of said correct answer category and unjust solution category differs.

[Claim 4] In the pattern recognition approach of recognizing the inputted pattern

using the pattern recognition dictionary drawn up by repeating an update process of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector The 1st formula which said update process imposes the 1st matrix on the 2nd vector which subtracted the 1st vector to said pattern vector, and was generated, and hangs transposition of the 2nd vector further, Said 1st vector from the characteristic vector of said 1st matrix The 2nd recurrence formula for updating the 1st recurrence formula and said 1st vector for updating said 1st matrix in the direction which optimizes the value of a valuation plan based on the recognition result of having followed the dissimilarity formula for calculating the dissimilarity of said pattern vector based on the subtracted vector is followed. The pattern recognition approach characterized by updating either [at least] said characteristic vector for every category, or the 1st vector for said every category.

[Claim 5] In the pattern recognition approach of recognizing the inputted pattern using the pattern recognition dictionary drawn up by repeating an update process of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector Said update process carries out the multiplication of the 1st multiplier to the square term of the inner product of said pattern vector and 1st vector. Furthermore, the 2nd multiplier The 1st formula to add The recurrence formula for updating said 1st vector in the direction which

optimizes the value of a valuation plan based on the recognition result of having followed the similarity formula for computing the similarity of said pattern vector to each category included at least is followed. Said 1st vector for every category

The pattern recognition approach characterized by updating.

[Claim 6] In the pattern recognition approach of recognizing the inputted pattern using the pattern recognition dictionary drawn up by repeating an update process of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector Said update process to the 1st valuation plan based on the recognition result of having followed the similarity formula with dictionary data with the limit beforehand determined as said pattern vector The pattern recognition approach characterized by updating said dictionary data for every category according to the recurrence formula for updating said dictionary data in the direction which optimizes the 2nd valuation plan adding what multiplied the formula showing the limit as which said dictionary data were furthermore determined beforehand by the constant of arbitration.

[Claim 7] In the pattern recognition approach of recognizing the inputted pattern using the pattern recognition dictionary drawn up by repeating an update process of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector Said update process updates said dictionary data for every category according to the recurrence formula for

updating said dictionary data in the direction which optimizes the valuation plan based on the recognition result of having followed the similarity formula of said pattern vector and dictionary data. With the function for determining the weight value which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category as said recurrence formula The difference of the similarity of said correct answer category and unjust solution category takes a high value in an equal part. The pattern recognition dictionary creation approach characterized by using said windowing function only when the difference of the similarity of said correct answer category and unjust solution category is predetermined within the limits in case it has the windowing function which takes a low value in the part from which the difference of the similarity of said correct answer category and unjust solution category differs and said dictionary data are updated.

[Claim 8] In the pattern recognition approach of recognizing the inputted pattern using the pattern recognition dictionary drawn up by repeating an update process of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector Said update process updates said dictionary data for every category according to the recurrence formula for updating said dictionary data in the direction which optimizes the valuation plan based on the recognition result of having followed the similarity formula of said

pattern vector and dictionary data. It has a windowing function for determining the weight value which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category as said recurrence formula. The parameter value which defines the property of this windowing function is the pattern recognition approach characterized by being set up by comparing two or more recognition results of said pattern vector.

[Claim 9] In the pattern recognition approach of recognizing the inputted pattern using the pattern recognition dictionary drawn up by repeating an update process of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector Said update process updates said dictionary data for every category according to the recurrence formula for updating said dictionary data in the direction which optimizes the valuation plan based on the recognition result of having followed the similarity formula of said pattern vector and dictionary data. It has a windowing function for determining the weight value which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category as said recurrence formula. The parameter value which defines the property of this windowing function is the pattern recognition approach characterized by what it opts for based on the frequency distribution about the

similarity of said correct answer category based on the recognition result of said pattern vector, and an unjust solution category.

[Claim 10] In the pattern recognition approach of recognizing the inputted pattern using the Oita pattern recognition dictionary drawn up by repeating an update process of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector Said update process The recurrence formula for updating said dictionary data in the direction which optimizes the valuation plan based on the recognition result of Oita of this pattern vector according to the similarity formula of said pattern vector and dictionary data is followed. Said dictionary data of a correct answer category and an unjust solution category The pattern recognition approach characterized by updating.

[Claim 11] It is the record medium which recorded the program for drawing up a pattern recognition dictionary by repeating renewal of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector and in which machine reading is possible. The recurrence formula which updates said matrix in the direction which optimizes the value of a valuation plan based on the recognition result of having followed the similarity formula for computing the similarity of this vector to each category containing the formula which imposes a matrix on said pattern vector and hangs transposition of this pattern vector further is followed. The record medium which recorded the

program which performs an updating means to make said matrix for every category update.

[Claim 12] The record medium according to claim 11 which memorized further the program which performs a calculation means to make the initial matrix of the matrix for said every category compute from the 1st matrix computed based on the set of said pattern vector, and the 2nd matrix for every category computed based on the set of said pattern vector classified into each category.

[Claim 13] It is the record medium which recorded the program for drawing up a pattern recognition dictionary by repeating renewal of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector and in which machine reading is possible. The 1st matrix is imposed on said pattern vector. Furthermore, transposition of this pattern vector The recurrence formula which updates said 1st matrix in the direction which optimizes the value of a valuation plan based on the recognition result of having followed the similarity formula for calculating the similarity of this pattern vector to each category based on the 2nd matrix which consists of a characteristic vector of the 1st formula to hang and said 1st matrix is followed. The program which performs an updating means to make said 1st matrix containing said characteristic vector for every category update is recorded. To said recurrence formula With the function for determining the weight value which becomes

settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category. The storage characterized by having the windowing function which takes a low value in the part from which the difference of the similarity of said correct answer category and unjust solution category takes a high value in an equal part, and the difference of the similarity of said correct answer category and unjust solution category differs.

[Claim 14] It is the record medium which recorded the program for drawing up a pattern recognition dictionary by repeating renewal of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector and in which machine reading is possible. The 1st formula which imposes the 1st matrix on the 2nd vector which subtracted the 1st vector to said pattern vector, and was generated, and hangs transposition of the 2nd vector further, Said 1st vector from the characteristic vector of said 1st matrix. The 2nd recurrence formula for updating the 1st recurrence formula and said 1st vector for updating said 1st matrix in the direction which optimizes the value of a valuation plan based on the recognition result of having followed the dissimilarity formula for calculating the dissimilarity of said pattern vector based on the subtracted vector is followed. The record medium which recorded the program which performs an updating means to make either [at least] said characteristic vector for every category, or the 1st vector for said every category update.

[Claim 15] It is the record medium which recorded the program for drawing up a pattern recognition dictionary by repeating renewal of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector and in which machine reading is possible. The multiplication of the 1st multiplier is carried out to the square term of the inner product of said pattern vector and 1st vector. Furthermore, the 2nd multiplier The 1st formula to add The recurrence formula for updating said 1st vector in the direction which optimizes the value of a valuation plan based on the recognition result of having followed the similarity formula for computing the similarity of said pattern vector to each category included at least is followed. Said 1st vector for every category The record medium which recorded the program which performs the updating means made to update.

[Claim 16] It is the record medium which recorded the program for drawing up a pattern recognition dictionary by repeating renewal of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector and in which machine reading is possible. To the 1st valuation plan based on the recognition result of having followed the similarity formula with dictionary data with the limit beforehand determined as said pattern vector The limit as which said dictionary data were furthermore determined beforehand The record medium which recorded the program which performs an updating means to

make said dictionary data for every category update in the direction which optimizes the 2nd valuation plan adding what multiplied the expressed formula by the constant of arbitration according to the recurrence formula for updating said dictionary data.

[Claim 17] It is the record medium which recorded the program for drawing up a pattern recognition dictionary by repeating renewal of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector and in which machine reading is possible. An updating means to make said dictionary data for every category update in the direction which optimizes the valuation plan based on the recognition result of having followed the similarity formula of said pattern vector and dictionary data according to the recurrence formula for updating said dictionary data, With the function for determining the weight value which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category, in case said dictionary data are updated The difference of the similarity of said correct answer category and unjust solution category takes a high value in an equal part. The record medium which recorded the program which performs a means to make the windowing function which takes a low value in the part from which the difference of the similarity of said correct answer category and unjust solution category differs apply to said recurrence formula

only when the difference of the similarity of said correct answer category and unjust solution category is predetermined within the limits.

[Claim 18] It is the record medium which recorded the program for drawing up a pattern recognition dictionary by repeating renewal of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector and in which machine reading is possible. An updating means to make said dictionary data for every category update in the direction which optimizes the valuation plan based on the recognition result of having followed the similarity formula of said pattern vector and dictionary data according to the recurrence formula for updating said dictionary data, The parameter value which defines the property of the windowing function for determining the weight value which said recurrence formula has, and which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category The record medium which recorded the program which performs the setting means made to set up by comparing two or more recognition results of said pattern vector.

[Claim 19] It is the record medium which recorded the program for drawing up a pattern recognition dictionary by repeating renewal of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector and in which machine reading is possible. An updating means to make

said dictionary data for every category update in the direction which optimizes the valuation plan based on the recognition result of having followed the similarity formula of said pattern vector and dictionary data according to the recurrence formula for updating said dictionary data, The parameter value which defines the property of the windowing function for determining the weight value which said recurrence formula has, and which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category The storage which recorded the program which performs the setting means made to set up based on the frequency distribution about the similarity of said correct answer category based on the recognition result of said pattern vector, and an unjust solution category.

[Claim 20] It is the record medium which recorded the program for drawing up a kind pattern recognition dictionary very much by repeating renewal of the dictionary data registered into the pattern recognition dictionary based on the inputted pattern vector and in which machine reading is possible. The recurrence formula for updating said dictionary data in the direction which optimizes the valuation plan based on the recognition result of Oita of this pattern vector according to the similarity formula of said pattern vector and dictionary data is followed. Said dictionary data of a correct answer category and an unjust solution category The storage which recorded the program which

performs the updating means made to update.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Field of the Invention] This invention relates to the dictionary creation approach in the pattern recognition field.

[0002]

[Description of the Prior Art] LVQ (study vector quantization) is introduced to reference "The Self-Organizing Map" (Proceedings of the IEEE (1990)). This sets an input vector to x , makes m a reference vector to each category, defines Dissimilarity s as $s=||x-m||$, and outputs the category which has the minimum dissimilarity in each category as a recognition result of an input vector x .

[0003] The following procedure is used for study of m at this time. It is the category which might be made into the recognition result when the training pattern x had been recognized as c and a reference vector m_c . At the time of a correct answer $m_c' = m_c + \alpha (x - m_c)$ -- (61)

It is $m_c' = m_c - \alpha (x - m_c)$ at the time of an unjust solution. -- (62)

** -- it is made like and a dictionary is updated. LVQ2 and LVQ3 are in approaches other than the above, and the formula (61) and the formula (62) are properly used based on the dissimilarity obtained as a recognition result.

[0004] Moreover, the method which applies the view called GPD to Euclidean distance is also learned. Let $R(x)$ be the recognition result obtained when x has been recognized. Generally as a multiplier decided by $R(x)$ in $\alpha(R(x))$, it is $mc' = mc + \alpha(R(x))(x - mc)$.

It considers as the updating Ruhr.

[0005] By these methods, dissimilarity is simple at Euclidean distance and the problem was in recognition precision. Moreover, it has also been a difficult technical problem to decide α the optimal, and there was a problem practically.

[0006] LSM (study subspace method) is introduced to reference "pattern recognition and a subspace method" (Sangyo Tosho Publishing (1986)). This sets an input vector to x , makes P a projection to each category, and is Similarity $s = x^T P x$. A definition is given. In addition, x^T It is the transposition vector of x . And a category with the maximum similarity in each category is outputted as a recognition result of x .

[0007] When the orthonormal vectors which constitute P are set to ps_{ii} , G shall be expressed as $P = GG^T$ by considering ps_{ii} as about the same poor matrix as r

piece. It carries out ψ_i as a recognition result obtained when x has been recognized, make $R(x)$ into $\psi_{ii}' = \psi_i + \alpha (R(x) - \psi_i)$, and ψ_i is orthonormalization of the ψ_{ii}' by Schmidt's method of orthogonalizing and is new. It learns by asking for ψ_{ii}' .

[0008] It is a format with mean-vector m , and is $s = \|x - m\|^2 - (x - m)^T P (x - m)$.

Although known as a Mahalanobis distance, it asks like ψ_i / ψ_{ii}' a top, and it is $m' = m + 2\alpha (R(x) - P(x - m))$ about m .

The method updated as be alike is also learned.

[0009] Moreover, the dictionary creation approach indicated by ALSM (average study subspace method) currently introduced to the same reference and Provisional Publication No. No. 137483 [56 to] is as follows.

[0010] P is constituted by making K into a positive value symmetric matrix using a part of characteristic vector of K , and similarity is defined as $s = x^T P x$. the case where the recognition of study is an unjust solution -- a correct answer category -- receiving -- a $K' = K + \alpha x x^T$ unjust solution category -- receiving -- K' -- it carries out by updating K by $K - \beta x x^T$.

[0011] By these approaches, what is going to optimize what theoretically might be indefinite, and the problem might be produced for recognition precision, without being unsuitable and study working.

[0012] On the other hand, the method with what [clear] it is going to optimize

theoretically is what applied GPD to the subspace method, and is introduced to reference "Discriminative Metric Design for Pattern Recognition" (ICASSP-95, Vol.5). This method expresses P of similarity $s=xTPx$ as a product showing two or more another rotations of Matrix Q_i , and it is constituted so that this Q_i may be learned. By this method, P and Q_i needed to be changed each time and there was a problem that learning time increased.

[0013]

[Problem(s) to be Solved by the Invention] This invention was made in view of the above-mentioned trouble, can take adjustment of the whole method theoretically in the dictionary creation approach in the pattern recognition field, and aims at offering the pattern recognition approach that pattern recognition with high-speed possible study and a high precision using the pattern recognition dictionary creation approach that dictionary creation with a high precision can be performed, and the dictionary drawn up by that cause can be performed.

[0014]

[Means for Solving the Problem] (1) Said update process used by the pattern recognition approach (claim 1, property nucleus study) of this invention Matrix K is imposed on the input pattern vector x. Furthermore, the transposition x^T of this input pattern vector The recurrence formula (equation (13)) which updates this

matrix K in the direction which optimizes the value J of a valuation plan (equation (6)) based on the recognition result of having followed similarity formula $s=xTKx$ for computing the similarity of this vector to each category containing the formula to hang is followed. The pattern recognition dictionary in which pattern recognition highly precise than updating said matrix K for every category is possible can create at a high speed.

[0015] (2) In said update process used by the pattern recognition approach (the creation approach of the initial matrix in property nucleus study of claim 2 and claim 1) of this invention The initial matrices (K0) of the matrix for said every category are the 1st matrix (Q) computed based on the set of said pattern vector, and a matrix computed from the 2nd matrix (P) for every category computed based on the set of said pattern vector classified into each category.

[0016] (3) Said update process used by the pattern recognition approach of this invention (claim 3, property nucleus study using ALSM: when using a window function) The 1st matrix (K) is imposed on said pattern vector (x). Furthermore, transposition of this pattern vector The value of a valuation plan (equation (6)) based on the recognition result of having followed the similarity formula (second= $xTPx$) for calculating the similarity of this pattern vector to each category based on the 2nd matrix (P) which consists of a characteristic vector (psii) of the 1st formula ($xTKx$) to hang and said 1st matrix Said 1st matrix which

contains said characteristic vector for every category according to the recurrence formula (formula (12)) which updates said 1st matrix in the direction to optimize is updated. To said recurrence formula With the function for determining the weight value which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category By the difference of the similarity of said correct answer category and unjust solution category taking a high value in an equal part, and having windowing-function $w(d)$ which takes a low value in the part from which the difference of the similarity of said correct answer category and unjust solution category differs, and (formula (8)) The pattern recognition dictionary in which highly precise pattern recognition is possible can create at a high speed.

[0017] (4) Said update process used by the pattern recognition approach (study of a constant and the mean vector is introduced into the property nucleus study using claim 4 and ALSM) of this invention The 1st formula which imposes the 1st matrix (K) on the 2nd vector $(x-m)$ which subtracted the 1st vector (m) to said pattern vector (x), and was generated, and hangs transposition of the 2nd vector further $(x-m)$ $(TK(x-m))$, The value J of a valuation plan (equation (6)) based on the recognition result of having followed the dissimilarity formula (equation (40)) for calculating the dissimilarity of said pattern vector based on the vector which subtracted said 1st vector from the characteristic vector (psii) of said 1st matrix

The 2nd recurrence formula (formula (42)) for updating the 1st recurrence formula (formula (41)) and said 1st vector for updating said 1st matrix in the direction to optimize is followed. By updating either [at least] said characteristic vector for every category, or the 1st vector for said every category, the pattern recognition dictionary in which highly precise pattern recognition is possible can create at a high speed.

[0018] (5) Said update process used by the pattern recognition approach (claim 5, thing which includes a multiplier and constant study by the study: formal [secondary] valuation plan type of a characteristic vector) of this invention The multiplication of the 1st multiplier (a_i) is carried out to the square term of the inner product of said pattern vector (x) and 1st vector (ps_{ii}). Furthermore, the 2nd multiplier (c) The 1st formula to add The similarity of said pattern vector included at least By updating said 2nd vector for every category according to the recurrence formula (equation (47)) for updating said 1st vector in the direction which optimizes the value of a valuation plan (equation (48)) based on the recognition result of having followed the similarity formula (an equation (45) - equation (46)) for calculating The pattern recognition dictionary in which highly precise pattern recognition is possible can create at a high speed.

[0019] (6) Said update process used by the pattern recognition approach (study of claim 6 and a characteristic vector: method of undetermined multipliers) of this

invention To the 1st valuation plan (the 1st term of an equation (53)) based on the recognition result of having followed the similarity formula (equation (52)) with dictionary data (normal direct vector ψ_{ii}) with the limit beforehand determined as said pattern vector The limit as which said dictionary data were furthermore determined beforehand By updating said dictionary data for every category according to the recurrence formula (formula (57)) for updating said dictionary data in the direction which optimizes the 2nd valuation plan (a formula (53) or formula (54)) adding what multiplied the expressed formula by the constant of arbitration (the 2nd term of a formula (53)) The pattern recognition dictionary in which highly precise pattern recognition is possible can create at a high speed.

[0020] (7) Said update process used by the pattern recognition approach (claim 7, window function) of this invention The recurrence formula for updating said dictionary data in the direction which optimizes the valuation plan (J) based on the recognition result of having followed the similarity formula (s) of said pattern vector and dictionary data (for example, K) According to (for example, formula (13)), said dictionary data for every category are updated. To said recurrence formula With the function for determining the weight value which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category The difference of the similarity of

said correct answer category and unjust solution category takes a high value in an equal part. In case it has windowing-function $w(d)$ which takes a low value in the part which is not so and said dictionary data are updated, only when the difference of the similarity of said correct answer category and unjust solution category is predetermined within the limits, the pattern recognition dictionary in which highly precise pattern recognition is possible can create at a high speed by using said windowing function.

[0021] (8) Said update process used by the pattern recognition approach (refer to claim 8, a watch locking-dog method, and drawing 4) of this invention According to the recurrence formula for updating said dictionary data, said dictionary data for every category are updated in the direction which optimizes the valuation plan based on the recognition result of having followed the similarity formula of said pattern vector and dictionary data. It has a windowing function for determining the weight value which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category as said recurrence formula. The parameter value which defines the property of this windowing function is set up by comparing two or more recognition results of said pattern vector. That is, the pattern recognition dictionary in which highly precise pattern recognition is possible can create at a high speed by checking the result of pattern recognition processing or the

recognition test by another test vector, resetting up the value of a parameter, in being equivalent to the situation which the recognition engine performance does not improve or worsens, returning required information, being made to carry out the restart of the study, and repeating this actuation.

[0022] (9) Said update process used by the pattern recognition approach (refer to claim 9, the thing using an adaptive window function, and drawing 6) of this invention According to the recurrence formula for updating said dictionary data, said dictionary data for every category are updated in the direction which optimizes the valuation plan based on the recognition result of having followed the similarity formula of said pattern vector and dictionary data. It has a windowing function for determining the weight value which becomes settled by the similarity of the correct answer category obtained as a recognition result at least, and an unjust solution category as said recurrence formula. The pattern recognition dictionary in which highly precise pattern recognition is possible can create the parameter value which defines the property of this windowing function at a high speed by being set up based on the frequency distribution about the similarity of said correct answer category based on the recognition result of said pattern vector, and an unjust solution category.

[0023] (10) Said update process used by the pattern recognition approach (claim 10, very kind correspondence) of this invention By updating said dictionary data

of a correct answer category and an unjust solution category according to the recurrence formula for updating said dictionary data in the direction which optimizes the valuation plan based on the recognition result of Oita of this pattern vector according to the similarity formula of said pattern vector and dictionary data The Oita pattern recognition dictionary in which the highly precise Oita pattern recognition is possible can create at a high speed.

[0024]

[Embodiment of the Invention] Hereafter, the operation gestalt of this invention is explained with reference to a drawing.

[0025] (1st operation gestalt) Drawing 1 is what showed the example of a configuration of the pattern recognition dictionary listing device concerning the 1st operation gestalt of this invention, and consists of the pattern input section 1 for inputting a training pattern, the recognition processing section 2, the renewal section 3 of a dictionary, the initial dictionary creation section 4, and the dictionary data storage section 5.

[0026] First, in the initial dictionary creation section 4, an initial dictionary is drawn up from the training pattern inputted from the pattern input section 1. When the drawn-up initial dictionary is stored in the dictionary data storage section 5 next, in the recognition processing section 2 Recognition processing is performed to the inputted training pattern using the initial dictionary stored in the

dictionary data storage section 5, or the dictionary in the middle of study, it is the renewal section 3 of a dictionary, and the recognition dictionary stored in the dictionary data storage section 5 based on the training pattern inputted as the result is updated. By repeating this updating until a terminating condition is fulfilled, study of a dictionary and amelioration are performed and a pattern recognition dictionary is drawn up.

[0027] Hereafter, processing actuation of the recognition processing section 2 and the renewal section 3 of a dictionary is mainly explained.

[0028] 1) Explain the renewal approach of a dictionary (property nucleus study) used for property nucleus study **** and the pattern recognition dictionary creation approach of this invention.

[0029] It is a transformation matrix and Similarity $s = xTKx$ -- (1) [as opposed to / set an input vector to x and / to each category / x for K]

A definition is given. The category which has the maximum similarity in each category about an input vector x shall be outputted as a recognition result of x .

[0030] Let the total number of patterns, and $p(x)$ into the appearance probability of x , and let $l(x)$ be a loss function for the transformation matrix corresponding to [in ω / corresponding to the category name of x for all training pattern sets and $C(x)$] the category name c for Kc , and M . Here, loss-function $l(x)$ is [Equation 1].

$$I(x) = 1/(1 + e^{-x}) \quad \dots (2)$$

[0031] It carries out. At this time, it is [Equation 2] about the valuation plan for performing evaluation to the recognition result of an input vector x.

$$J = \int_{x \in \Omega} I(s_{er}(x) - s_{ok}(x)) p(x) dx. \quad \dots (3)$$

$$s_{ok}(x) = \max_{C(x)=c} x^T K_c x = x^T K_{ok} x \quad \dots (4)$$

$$s_{er}(x) = \max_{C(x) \neq c} x^T K_c x = x^T K_{er} x. \quad \dots (5)$$

[0032] It carries out. Here, it is sok. The similarity to the transformation matrix corresponding to a recognition correct answer category when x has been recognized by the similarity formula (1) based on each transformation matrix K is supported, and it is ser. It is the similarity corresponding to an unjust solution category. In this case, it is considering as the category which shows maximum, respectively.

[0033] In addition, it is not the semantics of becoming a correct answer category with a correct answer as a recognition result but the semantics of the category which should serve as an answer of input data. The same is said of an unjust solution category.

[0034] Hereafter, in order to give explanation intelligible, it is I (ser(xk)-sok(xk))

about a loss function l .

Although expressed, even if it defines by $l(s_{er}(x_k), s_{ok}(x_k))$, a study method can be set up by the same view.

[0035] The transformation matrix corresponding to a correct answer category in Kok of a formula (4) and Ker are the transformation matrices corresponding to an unjust solution category.

[0036] Next, the case where the value J of a valuation plan (3) is minimized is considered. It is [Equation 3] when a formula (3) is expressed discretely.

$$J = \sum_{k=1}^M l(s_{er}(x_k) - s_{ok}(x_k)).$$

[0037] It becomes. It is [Equation 4] when this is differentiated with the element kij of K .

$$\frac{dJ}{dk_{ij}} = \sum_{k=1}^M l'(s_{er}(x_k) - s_{ok}(x_k)) \left(\frac{ds_{er}}{dk_{ij}} - \frac{ds_{ok}}{dk_{ij}} \right). \quad \dots (7)$$

ここで、

$$\omega(x) = l'(x) = \frac{e^{-x}}{(1+e^{-x})^2} = l(x)(1-l(x)). \quad \dots (8)$$

[0038] Moreover, $dk = s_{er}(x_k) - s_{ok}(x_k)$ It places and expresses $w(d) = l'(d)$. $w(d)$

is a function [****]. Since N is made into a number of dimension and Similarity s can be expressed like a formula (9), it is [Equation 5].

$$s = \sum_{i=1}^N \sum_{j=1}^N x_i x_j k_{ij} \quad \dots (9)$$

$$(k_{ij}) = K_{ok} \text{ の時、} \quad \frac{ds_{ok}}{dk_{ij}} = x_i x_j, \quad \frac{ds_{er}}{dk_{ij}} = 0. \quad \dots (10)$$

$$(k_{ij}) = K_{er} \text{ の時、} \quad \frac{ds_{ok}}{dk_{ij}} = 0, \quad \frac{ds_{er}}{dk_{ij}} = x_i x_j. \quad \dots (11)$$

[0039] It becomes. It is [Equation 6] in order to make J into min according to the steepest descent method.

$$k'_{ij} = k_{ij} - \alpha \frac{dJ}{dk_{ij}} = k_{ij} \pm \alpha \omega(d) x_i x_j, \quad \begin{cases} (k_{ij}) = K_{ok} \text{ (ok)} \\ (k_{ij}) = K_{er} \text{ (error)} \end{cases} \quad \dots (12)$$

[0040] Then, it is good. That is, the recurrence formula (study recurrence formula) for updating K can be expressed like a degree type (13).

[0041]

[Equation 7]

$$K' = K \pm \alpha \omega(d) x x^T, \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (13)$$

[0042] In addition, Matrix K is dictionary data stored in the dictionary data storage section 5, and Matrix K exists corresponding to each category, and it draws up a dictionary by updating this matrix K.

[0043] Next, with reference to the flow chart shown in drawing 3 , processing actuation of the pattern recognition dictionary listing device of drawing 1 is explained.

[0044] First, the initial matrix K_0 is created (after-mentioned). The variable i which shows each category is initialized to "0", and the following processings are repeated.

[0045] Step S3: If the input vector x for study is inputted, the matrix K_i for every category will be used for the dictionary data storage section 5, and similarity count will be performed.

[0046] Step S4: To the matrix K_i of the correct answer category which shows the maximum similarity, as are shown in a formula (13), and $\alpha w(d) xx^T$ is added and it is similarly shown in a formula (13) to the matrix K_i of the unjust solution category which shows the maximum similarity, subtract $\alpha w(d) xx^T$ and update Matrix K_i .

[0047] Processing of the above-mentioned step S3 - step S4 is performed about all categories until it increments step S5- step S6: i "1" every and fulfills a predetermined terminating condition.

[0048] You may make it make it end by a certain count, and may make it stop a terminating condition based on numeric data, such as a recognition rate, for example.

[0049] 2) In case the 2nd example dictionary of property nucleus study is updated (i.e., when creating K' from K) using a formula (13)), like ALSM shown as a well-known example, carry out at no times of a correct answer, but they receive a correct answer category at the time of an unjust solution. As opposed to a $K' = K + \alpha xx^T$ unjust solution category K can also be updated as $K' = K - \beta xx^T$. Here, α and β are the constants of arbitration.

[0050] although the conventional example of this approach is ALSM described previously -- ALSM -- the formula of Similarity $s = x^T P x$ it is -- similarity s used with the valuation plan (6) used to the thing in order to draw the study recurrence formula of this invention It was what offers the approach to dispel, and which made the mistake in being and was theoretically mistaken $s = x^T K x$. In this invention, right similarity count and a study recurrence formula are offered as it understands by the above-mentioned explanation. Thereby, a powerful recognition system can be constituted effectively.

[0051] Now, the vector m equivalent to the mean vector is introduced into the similarity formula (1) mentioned above and a study recurrence formula (13), and it is $s = (x - m)^T K (x - m)$ about dissimilarity. -- (14)

A definition can also be given. The study recurrence formula in this case can be expressed like a degree type (15) in consideration of ser and sok in I of a formula (13) interchanging.

[0052]

[Equation 8]

$$K' = K \mp \alpha \omega(d)(x - m)(x - m)^T, \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (15)$$

mは、

$$m' = m \pm 2\alpha \omega(d)K(x - m), \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (16)$$

で更新する。

[0053] In case it recognizes using this result, the similarity type using a part of characteristic vector obtained from Matrix K can be constituted, and it can also constitute so that this may recognize.

[0054] 3) The initial matrix K0 of step S1 of creation approach drawing 3 of the initial matrix K0 is the following, and can be made and searched for.

[0055] First, the set of all study vectors (the thing used for the above-mentioned serial study may also be different) is made into omega'all, and it is [Equation 9].

$$Q = \sum_{x \in \Omega_{all}} xx^T, \quad \dots (17)$$

[0056] The becoming covariance matrix Q is created and it asks for the characteristic vector. And matrix which compares the characteristic value corresponding to each characteristic vector, and consists of a characteristic vector with the large value T is made.

[0057] Next, the set of the study vector for every category is made into ω' , and it is [Equation 10].

$$Q' = \sum_{x \in \Omega'} (Tx)(Tx)^T, \quad \dots (18)$$

[0058] Becoming covariance-matrix Q' is created and the characteristic vector and characteristic value are calculated. and the diagonal matrix λ which arranged in the vertical angle the orthogonal matrix U and characteristic value which consist of a characteristic vector, and set others to "0" -- making -- $Q' = U^{-1} \lambda U$ -- suppose that it expresses. In the characteristic vector which constitutes this direct matrix U , the matrix P which consists of a corresponding thing which has large characteristic value is made, and initial transformation-matrix $K_0 = T P T^T P^T = (P^T)^T (P^T)$ is created.

[0059] Although KL conversion known well is used in the pattern recognition field as how to ask for T , as for the above-mentioned technique instead, the canonical distinction too known for the pattern recognition field well may be used. Namely, between-class covariance matrix S_b and the covariance matrix S_w in class are

used, and it is [Equation 11].

$$S_b \varphi = \lambda S_w \varphi, \quad \dots (19)$$

[0060] It is the approach of creating T from a ***** characteristic vector.

[0061] 4) the 3rd example of property nucleus study -- here, explain the case where a subspace method with weight is used as similarity.

[0062] Dissimilarity s is defined as $s = xTK^{-1}x$. K is a positive value symmetric matrix. Hereafter, it is written as $K^{-1} = H$. In this case, when a valuation plan (3) is expressed discretely, it comes to be shown in a formula (20).

[0063]

[Equation 12]

$$J = \sum_{k=1}^M l(s_{ok}(x_k) - s_{er}(x_k)). \quad \dots (20)$$

[0064] In addition, although it was similarity in the valuation plan (3), since it was dissimilarity here, the location of sok (xk) and ser (xk) was changed in the formula (20). If a formula (20) is differentiated with the element hij of H, a degree type (21) will be obtained.

[0065]

[Equation 13]

$$\frac{dJ}{dh_{ij}} = \sum_{k=1}^M l'(s_{ok}(x_k) - s_{er}(x_k)) \left(\frac{ds_{ok}}{dh_{ij}} - \frac{ds_{er}}{dh_{ij}} \right). \quad \dots (21)$$

ここで、 $\omega(d) = l'(s_{ok}(x_k) - s_{er}(x_k))$ と表す。

[0066] Moreover, Dissimilarity s is [Equation 14].

$$s = \sum_{i=1}^N \sum_{j=1}^N x_i x_j h_{ij} \quad \dots (22)$$

[0067] Since it is expressed, it is [Equation 15].

$$(h_{ij}) = H_{ok} \text{ の時、} \quad \frac{ds_{ok}}{dh_{ij}} = x_i x_j, \quad \frac{ds_{er}}{dh_{ij}} = 0. \quad \dots (23)$$

$$(h_{ij}) = H_{er} \text{ の時、} \quad \frac{ds_{ok}}{dh_{ij}} = 0, \quad \frac{ds_{er}}{dh_{ij}} = x_i x_j. \quad \dots (24)$$

[0068] *****. It is [Equation 16] in order to make J into min according to the steepest descent method.

$$h'_{ij} = h_{ij} - \alpha \frac{dJ}{dh_{ij}} = h_{ij} \mp \alpha \omega(d) x_i x_j. \quad \begin{cases} (h_{ij}) = H_{ok} \\ (h_{ij}) = H_{er} \end{cases} \quad \dots (25)$$

[0069] Then, it is good. That is, the recurrence formula for updating H can be expressed like a degree type (26).

[0070]

[Equation 17]

$$H' = H \mp \alpha \omega(d) x x^T, \quad \begin{cases} H = H_{ok} \\ H = H_{er} \end{cases} \quad \dots (26)$$

[0071] Here, it is referred to as $H' = H + \delta H$ and is $(H')^{-1} = K + \delta K$. It is [Equation 18] when it places.

$$I = (H + \delta H)(H + \delta H)^{-1} = (H + \delta H)(K + \delta K) =$$

$$I + \delta H \cdot K + H \cdot \delta K + \delta H \cdot \delta K \approx I + \delta H \cdot K + H \cdot \delta K$$

$$\delta H \cdot K = -H \cdot \delta K \quad K \cdot \delta H \cdot K = K \cdot H \cdot \delta K = -\delta K, \quad \delta K = -K \cdot \delta H \cdot K$$

[0072] On the other hand, it is [Equation 19].

$$\delta H = \mp \alpha \omega(d) x x^T$$

であるから

$$\delta K = \pm \alpha \omega(d) K x x^T K$$

$K^T = K$ なので

$$\delta K = \pm \alpha \omega(d) K x (K x)^T$$

K を更新する漸化式は

$$K' = K \pm \alpha \omega(d) K x (K x)^T \quad \dots (27)$$

となる。

[0073] In addition, it is the view of spherical-surface false BEIZU which the dimension close is performed as dissimilarity count and calculated as similarity by the following conversion in fact.

[0074]

[Equation 20]

$$s = \sum_{i=1}^N \frac{1}{\lambda_i} (x, \varphi_i)^2 \approx \sum_{i=1}^r \frac{1}{\lambda_i} (x, \varphi_i)^2 + \sum_{i=r+1}^N \frac{1}{h} (x, \varphi_i)^2 = \sum_{i=1}^r \left(\frac{1}{\lambda_i} - \frac{1}{h} \right) (x, \varphi_i)^2 + \frac{1}{h} \|x\|^2$$

$$= \frac{1}{h} \left[\|x\|^2 - \sum_{i=1}^r \left(1 - \frac{h}{\lambda_i} \right) (x, \varphi_i)^2 \right] \quad \dots (28)$$

[0075] In this case What is necessary is just to calculate the 2nd term of a formula (28) as similarity, since $\|x\|_2$ are fixed.

[0076] 5) the 4th example of property nucleus study -- here explains the case where the approach corresponding to BEIZU discernment is used.

[0077] It is [Equation 21] about dissimilarity.

$$s = (x - m)^T K^{-1} (x - m) \quad \dots (29)$$

[0078] A definition is given, and if the study recurrence formula for updating K is drawn like explanation of the 3rd renewal approach of a dictionary, a degree type (30) will be obtained.

[0079]

[Equation 22]

$$K' = K \pm \alpha \omega(d) K (x - m) (K (x - m))^T \quad \dots (30)$$

[0080] On the other hand, it is m. If it attaches, it asks for the recurrence formula

for updating m as follows. That is, a valuation plan is [Equation 23].

$$J = \sum_{k=1}^M I(s_{ok}(x_k) - s_{er}(x_k)). \quad \dots (3.1)$$

$$s = \sum_{i=1}^N \sum_{j=1}^N (x_i - m_i)(x_j - m_j)h_{ij}. \quad \dots (3.2)$$

$$\begin{aligned} \frac{dJ}{dm_p} &= \pm I'(d) \left\{ \sum_j (-1)(x_j - m_j)h_{pj} + \sum_i (x_i - m_i)(-1)h_{ip} \right\} \\ &= \mp 2I'(d) \sum_i h_{pi}(x_i - m_i) \quad \begin{cases} ok. \\ error. \end{cases} \quad \dots (3.3) \end{aligned}$$

ここで $h_{ij} = h_{ji}$ に注意。

$$\delta m = \pm 2I'(d)H(x - m).$$

$$m = m \pm 2\alpha\omega(d)H(x - m).$$

[0081] In false BEIZU, close count is performed as follows.

[0082]

[Equation 24]

$$H = \sum_{i=1}^N \frac{1}{\lambda_i} \varphi_i \varphi_i^T \quad \dots (34)$$

であるから、

$$\begin{aligned} H(x-m) &= \sum_{i=1}^N \frac{1}{\lambda_i} (x-m, \varphi_i) \varphi_i \approx \sum_{i=1}^r \frac{1}{\lambda_i} (x-m, \varphi_i) \varphi_i + \sum_{i=r+1}^N \frac{1}{h} (x-m, \varphi_i) \varphi_i \\ &= \frac{1}{h} \left\{ \sum_{i=1}^r \left(\frac{h}{\lambda_i} - 1 \right) (x-m, \varphi_i) \varphi_i + (x-m) \right\}. \quad \dots (35) \end{aligned}$$

よって、漸化式は以下のようになる。

$$m = m \pm \frac{2\alpha\omega(d)}{h} \left\{ (x-m) - \sum_{i=1}^h \left(1 - \frac{h}{\lambda_i} \right) (x-m, \varphi_i) \varphi_i \right\}. \quad \dots (36)$$

[0083] 6) Property nucleus study using ALSM : although loss-function differential was not used but alpha of a constant value and beta were used in the study recurrence formula in ALSM shown as a well-known example when a window function was used, this was unsuitable for learning exactly. It is desirable to use the windowing function which is equivalent to the differential of a loss function so that a study result may be exactly performed in the borderline of a correct answer and an unjust solution. Then, by this invention, windowing-function $w(d)$ which takes a value higher than the time of there being nothing right [that] is used in a part with equal similarity to a correct answer category and similarity to an unjust solution category, and it is [Equation 25] about a study recurrence

formula.

$$K' = K \pm \alpha \omega(d) x x^T. \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (37)$$

[0084] It comes out and constitutes. In addition, as a windowing function, a formula (7) can be used, for example.

[0085] Similarity here is $s = x^T P x$. P is a matrix which consists of a characteristic vector of K. It is [Equation 26] when this similarity is rewritten using characteristic vector psi of K.

$$s = \sum_{i=1}^r (x, \varphi_i)^2 (r < N) \quad \dots (38)$$

[0086] It becomes.

[0087] Approaches other than the above-mentioned ALSM type are also effective in this similarity formula (38). For example, [Equation 27] called false spherical-surface BEIZU

$$s = (x, \varphi)^2 + \sum_{i=1}^r \left(1 - \frac{\delta}{\lambda_i}\right) (x, \varphi_i)^2 + c. \quad \dots (39)$$

[0088] *****. Here, λ_{dai} is the characteristic value of K, and a scalar value it is decided from characteristic value that delta and c will be.

[0089] 7) Consider the case where the mean vector is further introduced into the property nucleus study using ALSM at the property nucleus study which used

installation and the above-mentioned ALSM for study of a constant and the mean vector. That is, the property nucleus study using ALSM which updates Matrix K and the mean vector as dictionary data stored in the dictionary data storage section 5 is explained.

[0090] The vector m equivalent to the mean vector is introduced into the similarity formula mentioned above or a recurrence formula, and it is projection distance, i.e., [Equation 28], about dissimilarity.

$$s = \|x - m\|^2 - \sum_{i=1}^r (x - m, \varphi_i)^2 \quad (r < N), \quad \dots \quad (4 \ 0)$$

[0091] A definition can also be given. The recurrence formula of the matrix K in this case is [Equation 29].

$$K' = K \mp \alpha \omega(d)(x - m)(x - m)^T, \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots \quad (4 \ 1)$$

[0092] It comes out, it is, and although m may actually calculate and ask for the mean vector from a study vector, it can also be updated with a recurrence formula as follows.

[0093]

[Equation 30]

$$m' = m \pm 2\alpha \omega(d)K(x - m). \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases}$$

[0094] (In this case, false BEIZU type dissimilarity, i.e., [Equation 31])

$$s = \|x - m\|^2 - \sum_{i=1}^r \left(1 - \frac{\delta}{\lambda_i}\right) (x - m, \varphi_i)^2 + c(r < N), \quad \dots (43)$$

[0095] ** -- it is effective.

[0096] The constant c contained in a formula (43) is [Equation 32].

$$c' = c \mp \alpha \omega(d), \quad \begin{cases} K = K_{ok} \\ K = K_{er} \end{cases} \quad \dots (44)$$

[0097] It can also update "Be alike."

[0098] 8) What includes a multiplier and constant study by the study (1) formal-

[secondary] valuation plan type of a characteristic vector (the norm may be contained)

Next, the case where learn characteristic vector psi of K serially directly, and renewal of a dictionary is performed is explained. That is, for example, it is [Equation 33] about similarity.

$$s = \sum_{i=0}^r a_i(x, \varphi_i)^2 + c \quad \cdots (45)$$

$$s = \sum_{i=0}^r a_i(x, \varphi_i)^2 + \sum_{i=0}^r b_i(x, \varphi_i) + c \quad \cdots (46)$$

$$s = \|x - m\|^2 - \sum_{i=0}^r a_i(x - m, \varphi_i)^2 + \sum_{i=0}^r b_i(x - m, \varphi_i) + c \quad \cdots (47)$$

[0099] ** -- the valuation plan based on [define by inner either and] this --

[Equation 34]

$$J = \sum_{x \in \Omega} l(s_{er}(x) - s_{ok}(x)) \quad \cdots (48)$$

[0100] It is the approach of determining better similarity and the dictionary data for dissimilarity count, by coming out and setting up and updating psi, m, ai, and bi serially in the direction which optimizes this valuation plan.

[0101] Since only psi and m were learned conventionally, if the recognition dictionary drawn up by this operation gestalt is used, as for pattern recognition capacity, precision will become high rather than the conventional example.

[0102] What is necessary is just to update psii and m with the same means as conventional LSM as a means to change serially, for example. Namely,

[Equation 35]

$$\varphi'_i = \varphi_i + \alpha(R(x))(x, \varphi_i)x \quad \dots (49)$$

$$m' = m + 2\alpha(R(x))((x - m) - \sum_{i=0}^r (\varphi_i, x - m)\varphi_i) \quad \dots (50)$$

[0103] The guarantee of a normal orthogonality of psii may use the approach which may use Schmidt's orthogonalization like the conventional example, and is shown in the 4th below-mentioned operation gestalt, and the other approaches are also possible for it.

[0104] About a_i , b_i , and c which were not in the conventional example, if a formula (47) defines similarity, it updates as follows.

[0105]

[Equation 36]

$$a'_i = a_i + \alpha(R(x))(x - m, \varphi_i)^2,$$

$$b'_i = b_i + \alpha(R(x))(x - m, \varphi_i),$$

$$c' = c + \alpha(R(x)),$$

[0106] (2) The recurrence formula of the study when using s (psii) for a general method of undetermined multipliers as the similarity based on orthonormal-vectors psii or dissimilarity had to save the normal orthogonality, and it is a problem whether what we do with it, and it had solved this using

Schmidt's method of orthogonalizing in LSM. In this invention, this is more theoretically transposed to the right approach. That is, it is a certain constant about a valuation plan. A is introduced and it sets up as follows.

[0107]

[Equation 37]

$$J = \sum_{x \in \Omega} \left\{ I(s_{er}(x) - s_{ok}(x)) + A \sum_i \sum_j (\delta_{ij} - (\varphi_i, \varphi_j))^2 \right\} \quad \dots (51)$$

$$\delta_{ij} = \begin{cases} 1(i = j). \\ 0(i \neq j). \end{cases}, \quad s_{ok}(x) = \max_{C(x)=c} s_c(x) = s_{ok}(x) \quad s_{er}(x) = \max_{C(x) \neq c} s_c(x) = s_{er}(x).$$

[0108] It learns by learning each parameter and a vector in the direction which optimizes this valuation plan. Hereafter, it explains concretely.

[0109] First, similarity is defined as follows.

[Equation 38]

$$s = \sum_{i=0}^r a_i(x, \varphi_i)^2 + c. \quad \dots (52)$$

[0110] A valuation plan is [Equation 39].

$$(\varphi_i, \varphi_j) = \begin{cases} 1(i = j). \\ 0(i \neq j). \end{cases}$$

[0111] It comes out, a coefficient A is introduced by the same technique as the

view of regularization in consideration of a certain thing, and it is [Equation 40].

$$J = \sum_{x \in \Omega} \left\{ l(s_{er}(x) - s_{ok}(x)) + A \sum_i \sum_j (\delta_{ij} - (\varphi_i, \varphi_j))^2 \right\} \quad \dots (53)$$

[0112] A valuation plan is set up.

[0113] It considers minimizing the value J of a formula (53). It is [Equation 41]

when a formula (53) is expressed discretely.

$$J = \sum_{k=1}^M l(s_{er}(x_k) - s_{ok}(x_k)) + A \sum_i \sum_j (\delta_{ij} - (\varphi_i, \varphi_j))^2 \quad \dots (54)$$

となる。ここで s は、

$$s = \sum_{i=0}^r a_i \left(\sum_{j=0}^n x_j \varphi_{ij} \right)^2 + c$$

であり、

$$X = A \sum_i \sum_j (\delta_{ij} - (\varphi_i, \varphi_j))^2$$

と置く。

$$\frac{ds}{d\varphi_{pq}} = a_p \cdot 2(x, \varphi_p) \cdot x_q, \quad \frac{ds}{d\varphi_p} = 2a_p (x, \varphi_p) x.$$

[0114] Supposing there is no term of X in a formula (54), the right-hand side of a

formula (54) is equivalent to the increment of a recurrence formula here, and a

recurrence formula is [Equation 42].

$$\varphi_p = \varphi_p \pm 2a_p \alpha \omega(d)(x, \varphi_p) x \quad \dots (55)$$

[0115] It becomes. On the other hand, it is [Equation 43].

$$\frac{ds}{d\alpha_p} = (\varphi_p, x)^2, \quad \frac{ds}{dc} = 1.$$

[0116] Next, it differentiates about X . X is a term which is not related to the similarity sok and ser to a correct answer category and an unjust solution category.

[0117]

[Equation 44]

$$\frac{dX}{d\varphi_{pq}} = A \sum_i \sum_j 2(\delta_{ij} - (\varphi_i, \varphi_j))' (\delta_{ij} - (\varphi_i, \varphi_j))'. \quad \dots (56)$$

$(\delta_{ij} - (\varphi_i, \varphi_j))'$ については φ_{pq} で微分すると、

$$i = p \text{ の時に } -(\varphi_{pq})' \cdot \varphi_{jp} = -\varphi_{jp},$$

$$j = p \text{ の時に } -\varphi_{iq} \cdot (\varphi_{pq})' = -\varphi_{iq},$$

のみが残る。従って、

$$\frac{dX}{d\varphi_{pq}} = A \sum_i \sum_j 2(\delta_{ij} - (\varphi_i, \varphi_j)) \{ -(\varphi_{pq})' \cdot \varphi_{jq} - \varphi_{iq} \cdot (\varphi_{pq})' \}$$

$$= 2A \left\{ \sum_j (\delta_{pj} - (\varphi_p, \varphi_j)) \cdot (-\varphi_{jq}) + \sum_i (\delta_{ip} - (\varphi_i, \varphi_p)) \cdot (-\varphi_{iq}) \right\}$$

$$= 2A \left\{ -\varphi_{pq} + \sum_j (\varphi_p, \varphi_j) \varphi_{jq} + -\varphi_{pq} + \sum_i (\varphi_i, \varphi_p) \varphi_{iq} \right\}.$$

$$\frac{dX}{d\varphi_p} = 2A - \varphi_p + \sum_j (\varphi_p, \varphi_j) \varphi_j + 2A - \varphi_p + \sum_i (\varphi_i, \varphi_p) \varphi_i$$

$$= 4A \left\{ \sum_{i=0}^r (\varphi_i, \varphi_p) \varphi_i - \varphi_p \right\}.$$

[0118] The recurrence formula of psi can be found as follows from these.

[0119]

[Equation 45]

$$\varphi_p = \varphi_p + \alpha \omega(d) \left[4A \left\{ \varphi_p - \sum_{i=0}^r (\varphi_i, \varphi_p) \varphi_i \right\} \pm 2a_p(x, \varphi_p)x \right] \begin{cases} ok. \\ error. \end{cases}$$

$$a_p = a_p \pm \alpha \omega(d) (\varphi_p, x)^2 \begin{cases} ok. \\ error. \end{cases}$$

$$c = c \pm \alpha \omega(d) \begin{cases} ok. \\ error. \end{cases} \quad \dots (57)$$

[0120] As mentioned above, it is [Equation 46] when deforming into a subspace method.

$$a_i \equiv 1, \quad c \equiv 1$$

[0121] It comes out, and it is and the recurrence formula of a formula (57) can be used as it is.

[0122] 9) In the study stated to the window function former, an update process is more accelerable by adding a device to windowing-function $w(d)$ used with a recurrence formula.

[0123] First, a high value is taken in $w(d)$ and the $d=0$ neighborhood, and the function (for example, $l'(x)$) which takes a low value on the outskirts is prepared.

On the other hand, although it is hardly helpful by controlling not to learn by preparing a certain threshold TH at the time of $w(d) < TH$, what the processing time requires is cut. This approach is or $(d < TH1)$ $(TH2 < d)$ further.

It is also controllable not to learn at the time of **. Of course, a value high [the $d=0$ neighborhood] as $w(d)$ is taken, a low value is taken on the outskirts, and it is the same on the outside of TH1 and TH2 using the function which serves as zero completely.

[0124] 10) It is similarity, when a study rule sets an input vector to x and sets dictionary data to λ at a general watch locking-dog method target $s=S(x, \lambda)$

A definition is given, $D(x, \lambda)$ of the study data decided from the windowing function $w(R(x))$ and input vector which are decided from data $R(x)$ obtained as a recognition result, dictionary data, etc. is used, and it is $\lambda' = \lambda + w(R(x), D(x, \lambda))$.

It can write. If w is not appropriately set up at this time, there is a trouble that study does not work, and a watch locking-dog method is adopted as an approach of solving this problem. That is, in case a watch locking-dog method here sets up the parameter value of a windowing function, when the result of the recognition performed in the case of study or the recognition test by the test data is checked, and the recognition engine performance does not improve or it worsens, it re(resetting up dictionary data etc. in that case) sets up so that the value of $w(R(x))$ may become small, and carries out the restart of the study.

With reference to the flow chart shown in drawing 4 , processing actuation of the

pattern recognition dictionary listing device in the case of adopting a watch locking-dog method as the setting approach of the parameter of a windowing function is explained.

[0125] First, recognition processing is performed using the dictionary drawn up beforehand (step S12). Next, it learns based on a recognition result (step S13). This is repeated N times (step S15 - step S16).

[0126] Next, a recognition result is compared with a former recognition result. However, when a comparison is the 1st time, the following is performed as that whose recognition precision is improving. That is, when getting worse than the time of recognition precision being before by evaluating a recognition processing result by step S17 or not going up, a dictionary, a parameter, etc. which had been progressed and saved to step S19 are returned. and it becomes a smaller value about $w(R(x))$ -- as -- setting up -- correcting (step S19 - step S20) -- step S11 -- return -- it performs again (step S21).

[0127] On the other hand, while [in step S17] recognition precision is improving as a result of evaluation, a current dictionary and a current parameter are saved (step S21), and return and activation are continued to step S11.

[0128] Study is terminated when an improvement is not accepted, even if it repeats this M times (step S21).

[0129] Here, the case where the differential of a degree type (58) is used is

taken for an example as a windowing function, and the setting approach of a parameter is explained more concretely.

[0130]

[Equation 47]

$$\sigma(x) = \frac{1}{1 + e^{-\alpha x}}, \quad \dots (58)$$

[0131] The facies of function sigma (x) shown in the formula (58) and the facies of thing sigma' (x) which differentiated it are shown in drawing 5 . If the value of alpha of a formula (58) becomes large at drawing 5 , signs that the value change of y to the value change of x becomes large, and the curve becomes steep in near the zero of sigma (x) are shown in (a) - (c).

[0132] As a loss function, the direction of a form, i.e., a step mold, when large is suitable for alpha. In any case, loss is the same at the time of right **, and it is because it is thought also at the time of misreading in any case that loss is the same. However, since differential becomes zero and stops learning in a perfect step function, the value of suitable alpha which is not infinite is needed as a windowing function.

[0133] So, better study comes to be performed by taking the more smallish value of alpha and enlarging at **** at the stage of the beginning of study. That is, it is the view which makes precise study perform to ****. What is necessary is just to

perform study in step S19 of drawing 4 , as the value of alpha of a formula (58) is enlarged in case the magnitude of $w(R(x))$ is changed.

[0134] 11) the thing using an adaptive window function -- when it generally says, recognition processing at the time of study is performed and the similarity to the category of a correct answer and the similarity to the category of an unjust solution are dangerous, it is desirable to take larger $w(d)$, and to take more smallish, when that is not right. Although it is appropriate to use the result since it exists in the place where 2 patterns of a correct answer and an unjust solution are near to the boundary line which divides a correct answer and an unjust solution when dangerous, it is because it is in the location where two patterns were widely different when that is not right, so using for study of a boundary line is unsuitable.

[0135] Then, $w(d)$ is set up with the following algorithms.

[0136] First, the difference d of the similarity value acquired from a recognition result is searched for about each training pattern, and the histogram about this is created. The example is shown in drawing 6 . The parameter v equivalent to the width of face of $w(d)$ is first decided from drawing 6 . As the arrangement direction, the following are mentioned, for example.

[0137] - Set the value of the one half of peak value P of distribution to T , and set the value on the axis of abscissa equivalent to the location, and the absolute

value of the difference of "0" to v.

[0138] - Calculate the average of the location which exists in a longitudinal direction about a right-hand side distribution part from $d=0$, and set it to v.

[0139] - The area from the location of $d=0$ to the location of v from $d=0$ to a right-hand side total distribution area The location which becomes the area which it is is set to v.

[0140] - Define v from the numeric value equivalent to the differential coefficient in $d=0$.

[0141] - Define v in the above-mentioned combination.

[0142] Next, for example, a degree type (59) defines windowing-function w (d) using this v.

[0143]

[Equation 48]

$$\omega(d) = \frac{e^{-\alpha d}}{(1 + e^{-\alpha d})^2}, \quad \alpha = D/v, \quad \dots (59)$$

[0144] D is the suitable multiplier defined beforehand here, and a formula (59) is the differential form of a formula (58). The windowing function shown in the formula (59) is carrying out Yamagata as shown in drawing 5, and if the value of v becomes small, it will change so that the width of face of a crest may become small. Therefore, if D is chosen appropriately, no matter what form the histogram

may have, it can set up so that $w(d)$ may have just right width of face.

[0145] study -- on the way -- for example, -- Once, the above-mentioned histogram is created in N time in the case of the similarity count for the study in step S13 of drawing 4, and it is made to perform study on and after next time in it using v which was able to be found there.

[0146] It is the process of study and structure which sets up v first in a case which seldom changes the width of face of distribution, and advances study by $w(d)$ based on it henceforth is also effective.

[0147] 12) Very much, although decision of the value of a windowing function and renewal of a dictionary are performed to the general one corresponding to a kind about the correct answer category and unjust solution category which take the maximum similarity, at the time of a kind, the option is very suitable. A kind is for being carried out in advance of detail recognition and narrowing down a recognition candidate very much.

[0148] Hereafter, it explains taking the case of the case where a recognition candidate is narrowed down to p pieces. In this case, since the correct answer category should be just contained by the p -th place, it is made to learn as follows.

[0149] Similarity of the category of the p -th place is set to ser , and it sets a dictionary to λ_{daer} , when the correct answer category is contained by the p -th place, and similarity of the category of the $p+1$ st place is set to ser , it sets a

dictionary to lambdaer and there is nothing right [that]. On the other hand, similarity of a correct answer category is set to sok, and a dictionary is set to lambdaok. It is w (ser, sok) about a windowing function. It carries out, $D(x, \lambda)$ of the study data called for from an input vector, dictionary data, etc. is used, and it is a recurrence formula about lambdaer and lambdaok. $\lambda' = \lambda + w(R(x)) D(x, \lambda)$

A dictionary is updated as be alike.

[0150] When the correct answer category is very contained in the kind candidate, you may control by the above-mentioned example not to update a dictionary, and you may make it use constant value as a windowing function. Moreover, you may control to perform renewal of a dictionary about all the unjust solution categories of the p-th less than place, and when the correct answer category is not very contained in the kind candidate, it may be made to update a dictionary from a correct answer category about the unjust solution category of a high order. For example, a dictionary is updated like the above by setting a dictionary to lambdaer in this case, setting similarity of an unjust solution category as ser.

[0151] (2nd operation gestalt) Drawing 2 is what showed the example of a configuration of the pattern recognition equipment concerning the 2nd operation gestalt of this invention, and recognizes an input configuration using the dictionary drawn up using the pattern recognition dictionary listing device as

shown in drawing 1 using the pattern recognition dictionary creation approach and it which were mentioned above. The dictionary drawn up by the above-mentioned dictionary creation approach is stored in the dictionary data storage section 5. In the recognition processing section 2 The dictionary data stored in the dictionary data storage section 5 are referred to to the pattern vector inputted through the pattern input section 1. Pattern recognition is performed (s is computed an input pattern vector, dictionary data, Similarity s, or whenever [similarity]), and a category with the largest similarity or a category with whenever [similarity / smallest] is outputted from the recognition result output section 6.

[0152] in addition, above-mentioned the 1- processing of drawing 1 indicated in the 2nd operation gestalt and each configuration section of drawing 2 can also be stored and distributed to record media, such as magnetic disks (a floppy disk, hard disk, etc.), optical disks (CD-ROM, DVD, etc.), and semiconductor memory, as a program which a computer can be made to execute.

[0153] Moreover, this invention is not limited to the above-mentioned operation gestalt, in the range which does not deviate from the summary, can deform variously and can be used.

[0154]

[Effect of the Invention] Since it uses the method which can be explained

theoretically that a recognition dictionary is updated in the suitable direction according to a recognition result according to this invention as explained above, the recognition dictionary changes in the direction in which recognition capacity is serially improved to the pattern data for study. Therefore, the pattern recognition dictionary in which highly precise pattern recognition is possible can be drawn up.

[0155] That is, since the study method which was able to take adjustment theoretically to the various dictionary study methods which had mismatching theoretically until now is offered, optimal suitable dictionary creation can be performed theoretically, and pattern recognition equipment highly precise as a result can be realized.

[0156] Moreover, it is effective also at the point which can perform dictionary creation in practical time amount by the device (for example, conversion with Matrix P and Matrix Q like GPD which are the conventional method is not performed) which learns at a high speed.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] Drawing having shown roughly the example of a configuration of the pattern recognition dictionary listing device concerning the 1st operation gestalt of this invention.

[Drawing 2] Drawing having shown roughly the example of a configuration of the pattern recognition equipment concerning the 2nd operation gestalt of this invention.

[Drawing 3] The flow chart for explaining processing actuation of the pattern recognition dictionary listing device of drawing 1 .

[Drawing 4] The flow chart for explaining processing actuation of the pattern recognition dictionary listing device in the case of adopting a watch locking-dog method as the setting approach of the parameter of a windowing function.

[Drawing 5] Drawing having shown change of the facies of the windowing function accompanying change of parameter value.

[Drawing 6] Drawing having shown an example of the frequency distribution of the difference of the similarity of a correct answer category and an unjust solution category based on the recognition result of a pattern vector.

[Description of Notations]

1 -- Pattern input section

2 -- Recognition processing section

3 -- Renewal section of a dictionary

4 -- Initial dictionary creation section

5 -- Dictionary data storage section

6 -- Recognition result output section